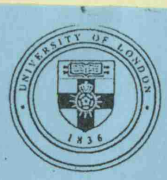


REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree Phd Year 2006 Name of Author WILLIAMS, A-J

Studies on Aspects that affect Chromatin Structure and Function of the Human CD2 Locus Control Region

Adam James Williams

Research thesis submitted for the degree of

Doctor of Philosophy at University College London

Supervisor: Dr Dimitris Kioussis

Division of Molecular Immunology

National Institute for Medical Research

Mill Hill, London

2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The overall aim of this project is to study the control of hCD2 gene expression and how different parts of the gene contribute to its regulation. This was achieved by searching for novel regulatory elements within the hCD2 gene whose function could be addressed by deleting them using an *in vivo* Cre/loxP deletion strategy. For this purpose it was necessary to develop tools that would, in future, enable the *in vivo* tissue-specific deletion of these identified sequences utilising this approach.

To identify novel regulatory elements within the hCD2 gene a number of different methods were used. Previously intron 4 has been shown to confer elevated levels of expression on hCD2 transgenes, however its function has never been studied in the context of variegation. Therefore, a construct including intron 4 and the truncated LCR was generated. This vector allowed elevated levels of hCD2 expression without conferring protection against PEV in transgenic mice, thus confirming that intron 4 could not compensate for a debilitated LCR. In addition, it was shown that decreasing transgene copy number whilst retaining the same site of integration affected hCD2 variegation in a manner that was dependant upon the site of integration. Therefore, we established that the inclusion of intron 4 in transgenes would be used for variegation studies whilst ensuring high levels of expression.

Using a bioinformatics approach a potential S/MAR site was identified within the 5' region of the hCD2 gene. DNase I hypersensitivity analysis revealed that the predicted S/MAR site did not form a DNase I HSS in transgenic mice. Furthermore, deletion analysis revealed that the predicted S/MAR site did not play an essential role in hCD2

expression in transgenic mice. In parallel, using RNase protection analysis and nested RT-PCR both sense and anti-sense transcripts originating from the hCD2 LCR were identified in transgenic mice. Finally, a differentially methylated region was identified within the hCD2 LCR, the position of which correlated with the presence of a previously identified DNase I HSS.

To enable the future tissue-specific deletion of potential regulatory elements from the hCD2 gene, two different Cre expressing transgenic lines were generated. Analysis revealed that the hCD2-iCre line mediated Cre recombination in almost all lymphoid cells, whereas, the Vav-iCre line mediated deletion in all haematopoietic cells that were analysed.

Acknowledgements

Firstly, I would like to thank Dimitris Kioussis for giving me the opportunity to work in his lab, and for his invaluable scientific guidance. The work presented in this thesis is entirely dependent on the mouse facility at NIMR and so I would like to thank Trisha, Keith, RoseMary and everyone in the animal houses for excellent animal husbandry and for all their help. I would like to thank Jasper de Boer and Eleni Ktistaki for their continuing advice on cloning and molecular biology. In addition, the generation and analysis of the Cre expressing transgenic mice was done in collaboration with Jasper de Boer. I would like to thank Nicky Harker for help the DNase I hypersensitivity assays. I would also like to thank Anna Garefalaki and Ursula Menzel for technical assistance in maintaining and manipulating ES cells in tissue culture. I would also like to thank Ursula Menzel for teaching me how to carry out RNase protection assays and Henrique Veiga-Fernandes for technical assistance with the ncRT-PCR. Finally, I would like to thank Amisha Patel for assisting with the cloning of constructs to generate Cre recombinase-inducible diphtheria knockin mice.

Contents

List of Figures and Tables

Abbreviations

AB IMDM	air buffered Iscove's modified Dulbecco's medium
APC	allophycocyanin
ddH_2O	double distilled H_2O
DNase I	deoxyribonuclease I
DN	double negative
DNMT	DNA methyltransferases
dNTP	deoxyribonucleotides
DP	double positive
ES	embryonic stem (cells)
EYFP	enhanced yellow fluorescent protein
FACS	fluorescence activated cell sorter
FCS	foetal calf serum
FITC	fluoresceine isocyanate
GFP	green fluorescent protein
HAT	histone acetyltransferase
HP1	heterochromatin protein 1
hCD2	human CD2
HDAC	histone deacetylase
HSC	haematopoietic stem cell
HSS	DNase I hypersensitive site
iCre	improved Cre recombinase
LCR	locus control region
LIF	leukaemia inhibitory factor

MFI	mean fluorescence intensity
PE	phycoerythrin
PEV	position effect variegation
S/MAR	scaffold/matrix attachment region
SP	single positive
TC	tricolour
TCR	T cell receptor
v/v	volume per volume
w/v	weight per volume

Publications

de Boer, J*., Williams, A*., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A. J. and Kioussis, D., Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur J Immunol* 2003. **33**: 314-325.

*These authors contributed equally to this work.

Chapter One

Introduction

1.1 Genome Structure and Organisation

1.1.1 Chromatin

The human nucleus is approximately 5-10μm in diameter but contains almost 2m of DNA. To fit this enormous length of DNA into such a small volume the DNA must be folded into a highly compacted state [1]. Importantly, this packaging must not prevent factors involved in transcription, recombination, repair and replication from gaining access to the DNA [1]. To achieve this level of compaction DNA is complexed with proteins to form chromatin.

Early light-microscopy revealed that there are two main types of chromatin called heterochromatin and euchromatin that are cytologically and functionally distinct [2]. Heterochromatin remains densely staining throughout the cell cycle, is late replicating, gene poor and is primarily composed of highly and moderately repetitive sequences [2]. Heterochromatin is generally considered to be repressive to gene transcription. In contrast, euchromatin is less condensed, appearing diffuse during interphase, and is largely early replicating [2, 3]. In general, actively transcribed genes are found within euchromatin. The properties of heterochromatin and euchromatin will be covered in more depth in section 1.8.

1.1.2 The Structure of Chromatin

Since its gross cytological characterisation the molecular structure of chromatin has been the subject of continued investigation. The fundamental subunit of all metazoan chromatin is the nucleosome which consists of approximately 146bp of DNA wrapped in a left handed superhelical turn approximately 1.75 times around an octamer of core

histone proteins (an $\{H3+H4\}_2$ tetramer and two {H2A+H2B} heterodimers) [4]. Each nucleosome is separated by 10-60 bp of linker DNA [5].

The core histones are among the most conserved proteins known which suggests that they play a fundamental role in nuclear physiology [5]. All of the core histones contain two domains that are separate both in structure and function. The globular domain, known as the 'histone fold' motif is responsible and sufficient for mediating histone-histone and histone-DNA contacts [6]. The amino-terminal and carboxy-terminal domains known as the 'histone tail' appear to have very little secondary structure, but are targets for a number of posttranslational modifications [6]. Even though the crystal structures of the histone tails remain unresolved they appear to emanate radially from the nucleosome, conveniently positioned to associate with linker DNA or adjacent nucleosomes [4, 6, 7]. This is supported by the observation that the histone tails appear to have very little significant impact in the structure and stability of individual nucleosomes, but play an important role in controlling the folding of nucleosomal arrays into higher ordered structures [8].

The nucleosomal array constitutes a chromatin fibre of ~11nm in diameter and is known as the 'beads on a string' arrangement [9]. Folding of this 11nm array generates a more condensed fibre of ~30nm (sometimes referred to as the 30nm solenoid) that is stabilised by the binding of linker histones such as H1 to the nucleosome core histones [6]. Interestingly, the linker histones are not related to the core histones in sequence, suggesting that they evolved independently of each other [10, 11].

The 30nm fibre is thought to be further folded within the nucleus to form 100-400nm thick interphase fibres, and then even further condensed to form the highly compacted metaphase chromosome structures [8].

It is important to note that the true structure of condensed chromatin fibres has yet to be determined in detail. In fact, the only structure to have been resolved convincingly is that of the nucleosomal core particle [4, 12].

1.1.3 Chromatin folding *in vivo*

The divalent cation concentration (2mM) required to drive 30nm fibre formation from simple nucleosomal arrays *in vitro* is far less than the estimated interphase nuclear divalent cation concentration (Ca^{2+} 4-6mM and Mg^{2+} 2-4mM) [6]. Thus, it is predicted that within the nucleus chromatin fibres would readily form heavily condensed structures [6]. This strongly suggests that extensive uncondensed 11nm nucleosomal arrays are unlikely to exist *in vivo*, and that the basic unit of chromatin competent for gene expression cannot be less than the 30nm fibre [6]. This observation is supported by a number of independent experimental systems. Electron microscopy studies of interphase chromosomes have revealed that they are mainly composed of 100nm-wide fibres, termed 'chromonema fibres' [13]. Following fibre decondensation during exit from mitosis the chromonema fibres were observed to first decondense to intermediated 100nm fibres and to 60nm fibres, and only occasionally were stretches of 30nm fibres observed [14].

In a subsequent study Tumbar et al. utilised a mammalian cell line containing an integrated array of LacI-binding sites to enable visualisation of the structure of actively

transcribed chromatin *in vivo* [15]. Using a LacI-GFP fusion protein they were able to decorate the 90-Mbp repeat tract in living cells [15]. The repeats were seen to form a single small dot within the nucleus. In contrast, when the LacI-GFP-VP16 transcriptional activation domain was used in the assay these foci were observed to rapidly decondense [15]. The decondensed tract appeared to be 80-100nm in diameter and coiled through a significant volume of the nucleus [15]. Bromouridine-5'-phosphate incorporation co-localised with the 100nm fibres suggesting active transcription was occurring within these compact fibres. Interestingly, RNA polymerase II (RNA Pol II) inhibitors did not affect the decondensation process, suggesting that this event occurred prior to the initiation of transcription [15]. One major criticism of this study is that it relies on a highly artificial system i.e. the LacI binding site repeats are massive and the VP16 transcriptional activation domain is extremely strong [6].

In a similar study Mueller et al. utilised a more natural system to examine the state of transcribed chromatin *in vivo* [16]. The system used a tandem array of the mouse mammary tumour virus promoter (MMTV) driving expression of a ras reporter gene [16]. To decorate the array in live cells they used a glucocorticoid receptor-GFP fusion protein [16]. Addition of a glucocorticoid receptor agonist led to the rapid decondensation of the array within 3 hours [16]. This rate was paralleled by an accumulation of ras mRNA [16]. Importantly, the decondensed transcriptionally active arrays were seen to form at least fully condensed 30nm fibres [16].

In both of these studies the 'decondensed' transcriptionally active sequences where seen to adopt at least a fully condensed 30nm chromatin fibre [6]. This suggests that the basic unit of uncondensed chromatin required to allow gene expression is the 30-100nm

fibre [6]. This is obviously difficult to reconcile with *in vitro* data where the presence of a single nucleosome is sufficient to inhibit RNA Pol II [17-19]. Furthermore, how the enormous RNA Pol II complex would penetrate such a condensed structure is not clear; the RNA Pol II initiation complex consists of more than 60 subunits and is over 3 mega Daltons, dwarfing the histone octamer at only 100 kilo Daltons [20].

1.2 Regulation of Chromatin Structure

The packaging of DNA into chromatin allows the genome to be efficiently compacted and stored within the nucleus. But packaging in this manner causes some obvious restrictions. First nucleotide sequences can become buried within the nucleosomal structure rendering them inaccessible or invisible to many regulatory factors [5]. In fact, only a few transcription factors can bind to sequences that are displayed on the surface of the nucleosome [21]. Secondly, during transcription RNA polymerase must process along the DNA as it scans the nucleotide sequence. The presence of nucleosomes on the DNA template would seemingly present a considerable obstacle to the polymerases. Thus, chromatin is inherently repressive to many processes that require the genome as a substrate. Therefore, to allow gene expression the chromatin structure must be remodelled by the decondensation of specific sequences to allow the transcriptional machinery access to the DNA template. Conversely, chromatin remodelling can actively prevent access to the DNA template and inhibit gene expression.

There are four main ways in which chromatin structure is regulated:

- Chemical modification of the histones
- Nucleosome repositioning or displacement

- Inclusion of histones variants
- DNA methylation

1.2.1 Chemical Modification of Histones and the Histone Code

Histones are subjected to a myriad of post-transcriptional modifications, including acetylation and methylation of lysines (K) and arginines (R), phosphorylation of serines (S) and threonines (T), ubiquitilation and sumoylation of lysines as well as ribosylation [8]. Furthermore, each lysine residue can be mono, di or even trimethylated, and arginines can be either mono or dimethylated [8].

The majority of these modifications occur on the amino-terminal and carboxy-terminal histone tail domains, although an ever-increasing number of modifications have been identified within the central globular histone domains [8]. These modifications can affect the nucleosome and chromatin structure in a number of ways. First, the addition of chemical moieties can alter the physical or chemical properties of the nucleosome/chromatin. [22] Second, these modifications may generate 'platforms' for the binding of regulatory proteins; or in the opposite manner may generate steric obstacles that preclude the binding of regulatory proteins [22]. Third, modifications within the globular domains frequently lie at the histone-DNA interface, suggesting a mechanism to directly disrupt the wrapping of DNA around the histone octamer [23].

Importantly, these modifications do not occur in isolation, which has two main implications. First, modification at a particular residue may preclude or enhance additional modifications on the same or alternative residues with the nucleosome [24]. Second, it is thought that specific combinations of these modifications act in concert to

direct particular biological events [25]. This is known as the 'histone code' hypothesis. Within the histone code, specific modifications can direct alternative functions depending upon the context in which they are found (the context in this case being the specific combination of posttranslational modifications that are present within the same or neighbouring nucleosomes) [25]. Histone modifications can also be maintained through cell division possibly encoding heritable epigenetic marks [26]. Although currently we do not have adequate understanding to decipher the histone code directly, gene specific and genome wide analysis in a number of model organisms has allowed the correlation of specific combinations of modifications with defined biological functions [27-35].

The last five years has seen an explosion in publications relating to histone modifications, and as a result it would be impossible to cover all the available data within the space of this introduction. Therefore only the best-characterised modifications will be covered in detail. Although the modifications will be discussed separately it is important to remember that that they co-exist within the nucleus.

1.2.2 Acetylation

Many histone residues have the potential to be acetylated *in vivo*, although the physiological significance of most are not well characterised. In general histone acetylation is regarded as having a positive influence on transcription [36]. However, the exact mechanisms by which acetylation influences transcription are not well defined. Interestingly, regions of hyperacetylated chromatin often exhibit increased general sensitivity to DNase I, suggesting addition of the acetyl moiety to the histone forces the chromatin to adopt a more accessible conformation *in vivo* [37]. However, the

precise level contributed by charge neutralisation (afforded by the removal of the positive charge on the lysine by acetylation) to the modulation of chromatin structure *in vivo* is difficult to determine [36]. Alternatively, the acetyl moiety may generate a platform for the recruitment of specific factors to the nucleosome [36].

Histones are acetylated by histone acetyl-transferases (HATs), which are generally classified as co-activators as they are unable to bind directly to DNA; instead they are recruited through interactions with independent DNA binding factors [22]. Many HATs are components of multi-subunit complexes such as the yeast SAGA and the human PCAF/GCN5 complexes [22]. Acetylation is a readily reversible process and there exists a large family of histone deacetylases (HDAC's) that exert a repressive function through the removal of histone acetyl moieties [38]. Within the nucleus an equilibrium between HATs and HDACs dictates the acetylation status of the genome, determining the expression state of each gene [39].

As mentioned above histone acetylation is generally regarded as having a positive influence on transcription [36]. For example, in yeast histone H3-K9/H3-K14 acetylation is enriched within the promoters and coding regions of actively transcribed genes, suggesting that acetylation is important for gene expression [28]. More recently, acetylation of lysine 56 within the globular domain of H3 (H3-K56) has been shown to be important for the expression of a sub-set of genes, through the recruitment of the SWI/SNF nucleosome remodelling complex [40]. Acetylation of histone H4, particularly on lysine 16 (H4-K16), has also been associated with transcription in yeast [33].

In contrast to yeast, human and mouse cells show enrichment of acetylated H3-K9/H3-K14 mainly in the 5' regions of active genes (including promoters, enhancers and locus control regions) and not within the coding regions [28, 29]. This difference can potentially be accounted for by the different gene sizes between the species; the average gene size in yeast is 2kb but in the human genome is approximately 27kb [28].

Although, histone hyperacetylation is generally associated with increased transcriptional activity, it appears that not all acetylation sites in the same histone molecule act as positive markers of gene expression [32]. For example, a study in yeast mapping global histone acetylation patterns revealed that as expected hyperacetylation of histone H3-K9/H3-K18/H3-K27 correlated with active transcription [32]. In contrast, the same areas of transcription displayed hypoacetylation of H4-K16 and H2B-K11/H2B-K16 [32]. This indicates that both acetylation and deacetylation of individual lysine residues may be involved in transcriptional activation. This may fit with the recent finding that direct deacetylation by the HDAC HOS2 is required for the expression of certain genes in yeast. [32]. In addition, histone acetylation may have important roles outside of the regulation of transcription. For example, acetylation of histone H4-K91 is associated with chromatin assembly *in vivo* [41].

1.2.3 Methylation

There are two types of histone methylation, targeting either arginine or lysine residues [22]. Histone arginine methylation is associated with gene activation [22]. Arginine specific histone methyltransferases (HMT) fall into two main families CARM1 and PRMT1 that target predominantly either H3 or H4, respectively [22]. Arginine HMTs

are classified as transcriptional coactivators as they are recruited to promoters indirectly through interactions with transcription factors [22].

In contrast, methylation of histone lysine residues can either be activating or repressive to transcription depending upon the residue methylated [42, 43]. The SET domain family member Suv39h1 methylates lysine 9 of histone H3 (H3-K9) [44]. This modification in turn recruits the silencing protein heterochromatin protein 1 (HP1). HP1 binds specifically to the methylated residues through its chromodomain driving the formation of heterochromatin [45, 46, Lachner, 2001 #880, 47]. Chromatin bound HP1 is able to recruit additional HMT activity thereby reinforcing and propagating the silenced state [48, 49]. Similarly, methylation of H3-K27 is able to recruit the silencing factor Polycomb, which in turn is able to recruit additional HMT activity [50]. Interestingly, methylation of H3-K9 does not always result in recruitment if HP1, in this case it appears that silencing is mediated through inhibition of the HAT complex p300 [51].

Methylation at lysine 4 of histone H3 (H3-K4) by the SET domain HMT SET1 (yeast) or SET9 (vertebrates) correlates with gene activation [52, 53]. Methylation of H3-K4 is able to prevent methylation of H3-K9, and also disrupts binding of the nucleosome remodelling and deacetylase (NuRD) repressor complex, potentially providing a mechanism by which this modification positively regulates transcription [54, 55]. Methylation of H3-K9 and H3-K4 therefore appear to mark regions of 'inactive' and 'active' chromatin, respectively [42, 43].

In yeast, methylation of H3-K4 is associated with the promoter and coding regions of actively transcribed genes [56]. Here it appears that dimethylation of H3-K4 marks active and activatable genes, whereas trimethylation of H3-K4 is an exclusive mark for highly transcribed genes [52, 56].

Similarly, genome-wide chromatin analysis in *Drosophila*, mouse and human has revealed that di/trimethylated H3-K4 and H3-K79 co-localise to regions of H3 hyperacetylation and are found almost exclusively within actively transcribed genes [27, 28, 30]. Interestingly, unlike in yeast, no differences have been observed in the distribution of dimethylated H3-K4 and trimethylated H3-K4 [27, 30].

Histone lysine methylation is thermodynamically an extremely stable modification [57]. Several mechanisms have been proposed for the removal of histone methylation, such as cleavage and removal of the histone N-termini, exchange with histone variants, or destabilisation by oxidation or by free radical attack [57]. Recently, Shi et al. have identified the first histone lysine demethylase, LSD1 (lysine-specific demethylase 1) [58]. LSD1 is highly specific for H3-K4, and none of the other lysine or arginine methylation sites in histone H3 and H4 can be demethylated by LSD1 [58]. As LSD1 acts as a co-repressor it is conceivable that it may be found in HDAC complexes [57].

It is also likely that additional demethylating enzymes will be identified. Indeed there are around 10 LSD1 related mammalian amine oxidases, six of which contain a putative nuclear localisation signal [57]. It is tempting to speculate that there are also activating histone lysine demethylases that are associated with HAT or nucleosomal remodelling activities to help revert repressive marks, such as H3-K9, H3-K27, and H4-K20

methylation [57]. It will also be interesting to discover if there exists an enzymatic mechanism to convert histone lysine trimethylation to an unmodified state [59]. One potential candidate for mediating this reaction is the hydroxylase-like protein Epe1 [59].

1.2.4 Phosphorylation

In vivo, histones can be phosphorylated on both serine and threonines residues [60]. Histone phosphorylation is involved in a number of diverse biological events, including transcriptional activation, transcriptional repression, facilitation of mitosis and meiosis, DNA repair, and apoptosis [60].

The phosphorylation of H3-S10 is important for condensation and segregation during mitosis and meiosis [61, 62]. Mitosis-specific phosphorylation of histone H3-S28 and H3-T11 also occurs, although the significance of these modifications is currently unknown [63, 64]. Histone H3 phosphorylation originates in pericentric heterochromatin and spreads throughout the genome during the G2-M transition of the cell cycle [65]. Histone H3 phosphorylation is essential of the initiation of chromosome condensation in mammalian and *T. thermophila* cells [61, 66]. Surprisingly, phosphorylation of histone H3-S10 is not required for cell cycle progression in yeast [67].

Phosphorylation of histone H3-S10 also has an important role in transcriptional activation. For example, transcriptional activation of the *c-fos* and *c-jun* immediate-early genes in fibroblasts is intimately associated with the phosphorylation of histone H3-S10 [68].

The manner in which histone H3 phosphorylation affects chromatin to mediate transcriptional activation is not defined. It is possible that the phosphorylation event causes an alteration in the physical properties of the nucleosome preventing the formation of higher ordered chromatin structures [60]. However, this hypothesis is difficult to reconcile with the role of this same modification in mediating chromosome condensation [60]. Interestingly, H3-S10 phosphorylation inhibits subsequent H3-K9 methylation, potentially providing a mechanism to maintain the active state [69].

Linker histones can also be phosphorylated [36]. This modification directly weakens the interaction of the basic tails of histone H1 with DNA [70]. Interestingly, phosphorylation appears to have even more influence on the interaction of H1 with chromatin, suggesting that it could result in the destabilisation of higher ordered folding of the chromatin fibres [70].

1.2.5 Ubiquitination

Ubiquitin is a 76 amino acid polypeptide that when added to proteins as a multimer marks the protein for degradation [71]. However, the core histones H2A, H2B, H3, the linker histone H1 and the histone variant H2A.Z all have the potential to be monoubiquitinated *in vivo*, a form of the modification that is not associated with protein degradation [72]. Histone ubiquitination is a readily reversible reaction, and like histone acetylation, ubiquitination is dynamic [73].

Ubiquitination of the C-terminal tail of H2B (lysine 123 in yeast and lysine 119 in vertebrates) is associated with transcriptional activation [72]. Ubiquitinated H2B is found to be concentrated at the 5' untranslated regions of highly transcribed genes in

flies and vertebrates [74-79]. In yeast, the absence of ubiquitinated H2B impairs the transcriptional activation of several highly inducible genes [80].

Little is known about how ubiquitination of histone mediates its effect on transcription. Although a relatively bulky modification, the presence of the ubiquitin moiety on H2B does not perturb nucleosome structure *per se*. Yet it is possible that ubiquitinated H2B serves as a platform for the binding of regulatory proteins through ubiquitin-interacting domains. Recent data have shown that ubiquitination of H2B selectively promotes the methylation of histone H3 in *trans*. For example, in yeast, mutation of the H2B ubiquitination site blocks methylation of histone H3-K4 and H3-K79 (both of which are associated with transcriptional activation) [81, 82]. Yet interestingly, H2B ubiquitination has also been implicated in mediating gene repression and gene silencing in yeast [83].

H2A can also be ubiquitinated, although the function of this modification is not clearly defined. Like ubiquitinated H2B, ubiquitinated H2A is incorporated into nucleosomes without any major structural disruptions [84]. H2A ubiquitination appears in some cases to be associated with transcriptional activity [79]. In contrast, other studies show that loss of ubiquitinated H2A correlates with increased transcription [85]. Ubiquitinated H2A has also been implicated in facilitating Polycomb mediated silencing [86]. In addition, ubiquitination of histone H2A has been shown to favour the binding of linker histones to nucleosomes, further connecting this modification to transcriptional silencing [87].

1.2.6 Nucleosome Repositioning and Displacement

The addition/removal of chemical moieties to histones cannot entirely explain how unrestricted access to the DNA template is achieved. One hypothesis is that chromatin is made fluid by chromatin-remodelling complexes that use the energy released from ATP-hydrolysis to continually shuffle the nucleosomes [88]. This process randomly (or non-randomly) exposes sequences that would otherwise be buried within the nucleosome, thereby rendering them accessible to regulatory factors. Thus in effect chromatin-remodelling complexes act to make chromatin 'transparent'. Yet this model is probably an oversimplification as chromatin-remodelling complexes can actually decrease accessibility to the DNA template through the generation of ordered nucleosomal arrays that facilitates folding into higher order condensed structures.

All ATP-dependent chromatin-remodelling factors identified so far are multi-subunit complexes that contain an ATPase subunit from the Swi2/Snf2 ATPase super-family [89]. There are four different classes of ATPase subunit, SWI/SNF, ISWI, CHD, and INO80, the presence of which determines the class of the chromatin-remodelling complex [1].

The ATP-dependent chromatin-remodelling factors have highly diverse biological roles *in vivo*, often with diametrically opposing consequences. For example, a single complex can affect transcription in either a positive or negative manner depending upon the context in which it is recruited. How can these differences be accounted for? The most-simple explanation can be gleamed from the understanding that each ATPase enzyme is embedded within distinct multi-subunit complexes [1]. The accessory proteins that

define these multi-subunit remodelling complexes are ultimately responsible for determining the consequence of the remodelling process.

1.2.7 The SWI/SNF Class

The founding member of the SWI/SNF class is the yeast SWI/SNF complex, which contains the Swi2/Snf2 ATPase subunit. In these mutants, transcription of the homothallic switching endonuclease (*HO*) gene and the sucrose-hydrolysing enzyme (*SUC2*) gene is reduced indicating that the SWI/SNF complex has a positive effect on transcription at these two loci [90, 91]. Consistent with this notion, SWI/SNF is required *in vivo* to generate an open chromatin domain at the *SUC2* gene promoter [92]. Furthermore, when *swi/snf*-mutant cells are grown in minimal media the transcription of many genes is reduced, suggesting that SWI/SNF complexes serve a global role in regulating inducible gene expression [93]. However, when the same cells are grown in rich media a number of genes actually show elevated levels of expression, indirectly suggesting a negative role for SWI/SNF in regulating transcription, at least at these loci [93]. With this in mind it is interesting to note that in mammalian cells HDAC proteins co-purify with SWI/SNF components [94, 95]. Finally, all SWI/SNF ATPases contain a bromodomain, which is a motif that interacts with acetylated histone tails, further suggesting that this family is involved in the positive regulation of transcription [1].

Yeast contains a second SWI/SNF complex RSC (remodels structure of chromatin) [1]. The catalytic subunit of RSC (Sth1) is absolutely required for cell viability [96]. The RSC complex appears to be involved in both positive and negative regulation of transcription [97, 98].

Like yeast, mammals contain two Swi2/Snf2 like ATPases, BRM and BRG1 (brahma and brahma related gene 1, respectively) [95]. These ATPases form distinct complexes with different biological and biochemical activities [95]. Mice lacking both *Brm* genes are viable, but show increased cell proliferation [99]. In contrast, *Brg1* mutant embryos die at the peri-implantation stage, and *Brg1* heterozygous mice are highly susceptible to developing tumours, suggesting an important role for mammalian SWI/SNF in cell cycle control [100]. With this in mind it is interesting to note that BRG1 physically associates with the tumour-suppressor proteins retinoblastoma and BRCA1 *in vivo* [101].

1.2.8 The ISWI Class

Drosophila ISWI (imitation switch) is the founding of the ISWI class of ATPases. ISWI-type ATPases are characterised by C-terminal SANT-like domains, which have been suggested to bind directly to histone N-terminal tails and stabilises them in a conformation that favours their binding to modifying enzymes [102, 103]. Although *Drosophila* contains only one ISWI homolog, it is found in three distinct complexes [1]. These are, NURF (nucleosome remodelling factor), ACF (ATP-utilising chromatin assembly and remodelling factor), and CHARAC (chromatin accessibility complex) [104-107].

Yeast has two members of the ISWI class of ATPases, Isw1 and Isw2 that form distinct complexes [108]. Both proteins repress transcription of early meiotic genes by recruitment to promoter region through interactions with the DNA-binding protein Ume6 [109-112]. Genome wide analysis of transcription indicates that Isw2 complexes may be involved in the repression of genes other than the early meiotic genes, also by

the generation of closed chromatin [113]. These results imply that the ISWI and SWI/SNF complexes may have opposing roles in regulating chromatin accessibility [1]. However, Isw1 appears to have diverse functions and has been implicated in ensuring co-ordination of elongation and efficient termination of transcription by RNA Pol II [114].

Like yeast, mammals contain two ISWI homologues, SNF2H and SNF2L. *In vivo*, SNF2H is ubiquitously expressed, with highest expression in populations of actively dividing cells [115]. SNF2H is absolutely required for survival and *Snf2h*$^{-/-}$ embryos die during the peri-implantation stage, resulting from growth arrest and cell death in both the trophectoderm and inner cell mass [116]. In contrast, SNF2L is mainly expressed in brain tissue, where it forms two remodelling complexes [115, 117, 118]. SNF2L is a component of the human NURF complex, which drives transcriptional activation of the engrailed genes during neuronal development [117]. SNF2L is also found within the novel chromatin remodelling complex CERF (CECR2-containing remodelling factor) [118]. An *Snf2l* knockout has not yet be described.

1.2.9 The CHD Class

The CHD class of ATPases is characterised by the presence of a chromodomain, a motif that can mediate binding to methylated histones [1]. *Drosophila* Chd1 (a CHD class ATPase) protein has been localised to interbands and puffs in polytene chromosomes, indicating a positive role in transcription [119]. Furthermore, through one of its two chromodomains, Chd1 is able to bind to methylated H3-K4 (a modification enriched in actively transcribed chromatin), and recruit the HAT complex SLIK to enhance nucleosomal acetylation [120]. However, Chd1 appears to have diverse biological roles.

For example, Chd1 has been shown to function as an ATP-utilising chromatin assembly factor, catalysing the transfer of histones from the NAP1 chaperone to DNA by a processive mechanism yielding regularly spaced nucleosomes [121].

The mammalian CHD ATPase, Mi-2, is found in the NuRD complex (nucleosome remodelling deacetylase), again suggesting a role in transcriptional repression [122]. Mi-2 complexes also contain putative methylated-DNA-binding proteins, providing a potential link between DNA methylation and chromatin remodelling [123]. Indeed, Mi-2 complexes function more efficiently on chromatin templates containing methylated DNA [123]. However, the function of Mi-2 appears to be context dependent. For example, during thymocyte development Mi-2 interacts with the HAT protein p300 to promote hyperacetylation of the CD4 enhancer and allow expression [124].

In yeast, mutations in *chd1* have a small impact on global transcription with only 2-4% genes affected >2-fold [125]. Yeast Chd1 has been shown to interact with transcription elongation factors and localises to actively transcribed genes, implying a role in controlling transcription elongation through chromatin remodelling [126].

1.2.10 The INO80 Class

The INO80 class is characterised by the presence of an insert within their ATPase domain, splitting it into two separate parts [1]. The ISWI-like ATPase Ino80 was originally identified in yeast as a gene essential for the transcriptional activation of the *INO1* gene that is induced in the absence of inositol. SWR1P is another member of this class [127].

1.2.11 Intrinsic Nucleosome Dynamics

Before discussing the mechanisms employed by chromatin remodelling enzymes it is worthwhile to highlight a number of intrinsic physical properties of nucleosomes.

DNA is a moderately flexible polymer with a persistence length of about 150bp [128]. This means that in the absence of exogenous forces DNA will follow a relatively straight path, but when present in nucleosomes it is twisted in 1.65 toroidal superhelical turns around the histone octamer and thus is severely distorted [129]. This means bending DNA around the histone octamer occurs at high-energy costs, which is compensated by DNA-histone interactions that occur every 10bp on each DNA strand, generating 7 histone-DNA interactions clusters per DNA coil [4, 130, 131]. These DNA-histone interactions are stabilised by approximately 116 direct and 358 water-bridged interactions rendering the nucleosomal particle stable under physiological conditions [132, 133]. However, due to the energy costs expended in twisting the DNA around the octamer, nucleosomes are far less stable than the total number of DNA-histone interactions would imply [134]. Furthermore, the histone octamer itself is unstable when not complexed with DNA [135]. Taken together these properties indicate that the nucleosome is relatively unstable and as such a potentially dynamic entity [133]. This prediction is supported by the observation that passive DNA-binding proteins and restriction enzymes can often bind to their target sites even when they buried within nucleosomes or polynucleosomal arrays [21, 136-139]. The mechanisms and conformational changes underlying such spontaneous (uncatalysed) site accessibility are not yet defined in detail. However, one early study concluded that nucleosomes have an inherent ability to slide or translocate along the DNA template [140]. Yet, experiments to directly test this conclusion have demonstrated that this is

unlikely to be the main mechanisms responsible for mediating spontaneous accessibility [141]. More recent studies have revealed that under physiological conditions nucleosomes undergo spontaneous large-scale conformational changes that involve the transient unwrapping of the histone bound DNA, making all regions of the nucleosomal DNA ephemerally accessible [142]. These spontaneous fluctuations in nucleosomal structure transiently expose 'buried' DNA sequences, allowing access to DNA binding factors that would otherwise be occluded [142]. Furthermore, nucleosomes shift their conformational equilibrium in response to the presence of exogenous site-specific DNA-binding proteins, appearing to facilitate their own 'invasion' [142]. Interestingly, when a site-specific factor binds to nucleosomal DNA the histone octamer does not slide out of the way, but rather appears to remain fixed in place along the remaining wrapped DNA [142]. Thus, it seems that nucleosomes (at least at the level of the simple array) do not present an impermeable barrier to exogenous DNA-binding factors.

Two main mechanisms have been proposed to explain nucleosome sliding and spontaneous accessibility; twist diffusion and loop/bulge diffusion [134].

In twist diffusion Brownian energy fluctuations would twist the DNA helix at the edge of the nucleosome, leading to the replacement of the canonical histone-DNA interactions with equivalent ones on neighbouring base pairs [134]. Propagation of this distortion around the surface of the histone octamer would change both the rotational and translational position of the nucleosome as the DNA 'screws' over its surface [134]. Because of the high torsional flexibility of DNA not all the histone-DNA contacts would need to be broken simultaneously, consequently reducing the thermodynamic barrier to twisting [134]. However, the observation that nucleosomes frequently

maintain their rotational positioning argues against this as a principle sliding mechanism [135].

In loop or bulge diffusion, thermal energy fluctuations would result in the detachment of DNA at the point of entry/exit from the nucleosome [135]. Dissociation of the first 30-35 bp of DNA from either end requires relatively little energy (whereas dissociation of more internal contacts requires increasing amounts of energy) [134]. The detached segment of DNA may then either rebind to form the original nucleosome, or alternatively may be replaced by a more distal sequence of DNA, thereby creating a DNA loop or bulge on the nucleosome surface. The directional propagation of this loop around the histone octamer would alter the translational (but not necessarily the rotational) position of the nucleosome [134]. The distance that the nucleosome slides would be dictated by the size of the loop [134]. Loop propagation requires minimal input energy as each histone-DNA contact that is broken ahead of the loop is immediately replaced by a histone-DNA contact behind the loop [134]. In this model detached DNA within the loop may be exposed to interactions with transcription factors, thereby explaining the phenomenon of spontaneous accessibility. Importantly, this model allows for more than simple translational repositioning of the nucleosome. For example, interaction of the exposed region of the histone octamer with an unconnected region of DNA could result in the displacement of the octamer in *trans* [134]. Alternatively the displaced region of DNA could interact with a second histone octamer resulting in the formation of a di-nucleosome structure [134].

1.2.12 Mechanisms of ATP-Dependent Chromatin Remodelling

The Snf2p-related enzymatic sub-unit of all ATP-dependant chromatin-remodelling complexes shares sequence homology with the large helicase superfamily 2 (SF2) [89]. The SF2 family is a diverse group of proteins that include true helicases that drive strand separation, as well as ATP-dependent DNA translocases [89]. This led to the prediction that the Snf2p-related core proteins of the remodelling complexes would function as ATP-dependant DNA translocases [89]. Indeed, direct evidence for this hypothesis is now available, although translocation by these factors does not seem to be highly processive [143-145]. How this DNA translocation is exploited to drive chromatin remodelling is not yet defined. However, it has been suggested that nucleosome-remodelling enzymes simply enhance the intrinsic dynamic properties of nucleosomes by lowering energy barriers through destabilisation of histone-DNA interactions [134]. Alternatively, they could act as Brownian ratchets to capture the large-scale DNA transitions that occur spontaneously under physiological conditions.

As for sliding and spontaneous accessibility, both twisting and looping models have been invoked to explain the action of remodelling complexes. However, inclusion of barriers to twist diffusion, such as nicks or hairpin structures, does not prevent remodelling [143, 146-149]. Indeed, most experimental data are more consistent with the bulge diffusion model, nonetheless, they do not rule out a contribution of twist diffusion in nucleosome remodelling *in vivo*.

Remodelling can have outcomes other than the simple sliding of nucleosomes along the DNA strand in *cis*, such as histone H2A/H2B dimer removal or replacement [89]. Histone octamers are unstable at physiological salt concentrations and DNA is required

for the stable association of H2A/H2B dimers within nucleosomes [89]. Displacement, by chromatin remodelling complexes, of up to 50 bp of nucleosomal DNA is expected to destabilise the nucleosome enough to result in the loss of one H2A/H2B dimer [89]. Dimer removal could prevent the formation of ordered nucleosomal arrays thereby inhibiting folding into higher ordered structures. In an extreme case dimer removal may be the first step in the complete removal of histone octamers from the DNA, as has been suggested for the SWI/SNF complex removal of nucleosomes from the *PHO5* promoter *in vivo* [150, 151]. Interestingly, disruption of nucleosome structure is not absolutely required for remodelling *per se* as cross-linking histones within the nucleosome does not affect nucleosome sliding *in vitro* [152-154].

Histone replacement is another important mechanism to remodel chromatin through alterations in nucleosome composition. The SWR1 chromatin remodelling complex is able to catalyse the ATP-dependant exchange of H2A with the histone variant H2A.Z in a replication independent manner [155]. Interestingly, the histone variant H2A.Z has been shown to protect euchromatin from the ectopic spreading of silent heterochromatin [156].

In conclusion, the bulge diffusion model can explain almost all phenomena displayed *in vitro* by the chromatin remodelling complexes [134]. It also appears that the apparent differences between remodelling classes can be attributed to the size of loop that the enzyme creates on the surface of the octamer [134].

1.2.13 Nucleosomes, RNA Polymerase II and Chromatin Remodelling

In vivo, RNA Pol II is able to transcribe at a rate of 25 nucleotides per second [157]. In contrast, transcription *in vitro* can be inhibited by the presence of a single nucleosome [17-19]. How does the polymerase achieve such high rates of transcription *in vivo*? One possibility is that nucleosomes are entirely depleted from actively transcribed genes. However, this is an unlikely event, indeed, the CHIP assay, which is one of the fundamental techniques now used to study gene expression, is based on the detection of covalently modified histones positioned within actively transcribed genes. Therefore, there must exist additional mechanisms *in vivo* to allow transcription to proceed unhindered through nucleosome bound sequences. As mentioned above RNA Pol II is unable to transcribe through nucleosomes *in vitro*. However, when salt concentrations are raised sufficiently for one dimer to be displaced (leaving behind a histone hexamer), transcription can proceed albeit at a slow pace [158]. Surprisingly, the position of the remaining histone hexamer is unaltered after transit of the polymerase [158]. How the polymerase actually traverses the hexamer is not known.

An important characteristic of Pol II is that *in vivo* it functions in cooperation with a number of accessory complexes such as FACT (facilitates active transcription) [159]. When FACT is added to *in vitro* transcription assays it is able to facilitate elongation of RNA Pol II through nucleosome bound templates even at low salt concentrations [159, 160]. Once again elongation is accompanied by the displacement of a single dimer from each nucleosome [160]. The current hypothesis is that *in vivo* FACT displaces histone H2A/H2B dimers from nucleosomes ahead of the polymerase, to allow its transit, and then possibly reassembles them behind it [161]. In support of such a mechanism,

chromatin *in vivo* containing transcribed genes, has been shown to be deficient in histones H2A and H2B [162].

This hypothesis suggests that the transcriptional machinery is able to disrupt the histone octamer directly. This could potentially facilitate subsequent transcription in a number of ways. First, the loss of H2A/H2B dimers could decrease the ability of the locus to fold into higher-ordered structures [158]. Second, it is possible that the transiently disrupted chromatin left behind the transcribing Pol II may serve as a window of opportunity for the binding of factors that further destabilise nucleosome structure [158]. Indeed, depletion of H2A/H2B facilitates the binding of transcription factors to nucleosome-covered promoters *in vitro* [163].

In addition to directly disrupting chromatin structure, mounting evidence suggests that the RNA Pol II may also direct posttranscriptional histone modifications [164]. Indeed, certain nucleosomal modifications are only found at actively transcribed genes [30]. It has been suggested that that the polymerase C-terminal domain (CTD) serves as a platform for the binding of chromatin modifying complexes [165]. For example, in human cells the HAT PCAF has been shown to bind specifically to the phosphorylated CTD of RNA Pol II [165]. It is thought that the associated PCAF complex is able to acetylate histones as the polymerase transcribes through the locus [165]. Similarly, in yeast, histone methylation appears to be targeted to the 5' ends of genes via an interaction between the phosphorylated CTD of Pol II and a methylase-containing complex [52, 166].

1.2.14 Nucleosome heterogeneity and histone variants

In vivo, chromatin is an extremely heterogeneous nucleoprotein even at the nucleosomal level. There exist a number of histone variants each with different physical and chemical properties (for a recent review see [167]). It is thought that incorporation of these variants into specific regions creates domains of chromatin with novel properties. However, in most cases it is not known how the histone variants affect nucleosome structure and chromatin folding. Nor is it understood how many of these variants are localised to specific DNA sequences.

1.2.15 DNA Methylation

DNA methylation occurs on cytosine residues within a number of species. In mammals, DNA can be methylated on cytosines within short CpG repeats, while in plants, cytosine methylation is associated with three different nucleotide sequence contexts CG, CNG and CNN (where N is A, T or G) [168]. Interestingly, there is little or no DNA methylation in the genomes of yeast, flies or nematodes [169].

In both mammals and plants DNA methylation is associated with gene silencing, which is mediated through a number of different mechanisms. The first is steric occlusion of DNA binding proteins. For example, in mammals, DNA methylation within the binding sequence of the transcriptional repressor protein CTCF is able to prevent its binding. Second, proteins that bind specifically to methylated DNA sequences can recruit additional regulatory proteins. The mammalian methyl-CpG binding protein MeCP2 is able to recruit the ATP-dependent chromatin remodelling complex Brm to methylated DNA and mediate gene silencing [170]. Third, recent evidence provides a direct link between DNA methylation and histone modification. For example, in mammals, the

DNA-methyltransferases DNMT1 and DNMT3b interact with the histone H3-K9 HMT SUV39H and HP1, respectively [171-173]. Furthermore, the loss of methylated H3-K9 in *Suv39h*$^{-/-}$ knockout embryonic stem cells results in a reduction of DNA methylation at centromeric satellite repeats [171]. Conversely, treating cells with a DNA methylation inhibitor results in a reduction of methylated H3-K9 [174]. Thus, it appears that there is a bi-directional dependence between DNA and histone methylation. Similar associations have been described in *Neurospora crassa* and *Arabidopsis* [175-177].

DNA methylation has been shown to be a late event in gene silencing in a number of systems, indicating that it may not be a primary mechanism used to initiate gene silencing [178, 179]. However, DNA methylation is epigenetically stable as hemimethylated sites generated during DNA replication are recognised and methylated by maintenance methyltransferases, such as mammalian DNMT1 [180, 181]. This suggests that DNA methylation may instead serve as an epigenetic mark for maintaining silent chromatin [178].

1.3 The Structure of Mitotic Chromosomes

During interphase, chromosomes are relatively uncondensed to allow transcription. As cells enter mitosis the chromosomes undergo condensation to form small rod shaped structures. This condensation allows sister chromatids to be separated and transported to opposite poles of the dividing cell. Although mitotic chromosomes can be readily visualised and even isolated as compact bodies, their structure has yet to be resolved. In part this is due to the high density of metaphase chromosomes, which makes viewing the internal structures virtually impossible [182]. In addition, chromatin is extremely

sensitive to small changes in ionic conditions, meaning that different preparation methods result in different structural models [182].

Early experiments used high salt extraction of chromosomes to remove histone proteins and enable visualisation of internal structure [183, 184]. Chromosomes treated in this manner were seen to contain a proteinaceous core, approximately the same size and shape as the original chromosome, with loops of DNA extending from the central protein core [183, 184]. The use of nucleases during scaffold isolation revealed that scaffolds were stable in the absence of DNA, suggesting that scaffolds are structurally independent entities and that DNA plays no structural role [184]. This gave rise to the simple 'radial loop model' in which a central protein scaffold defines the structure of the chromosome and organises the DNA into ~50kb loops [183, 184].

One of the most profitable lines of investigation stemming from the establishment of the radial loop model has been the identification and characterisation of scaffold components. Two major components of the scaffold, originally named SCI and SCII, were later found to be topoisomerase II and an SMC subunit of the condensing complex respectively [185-188]. In isolated, intact chromosomes both of these proteins showed an axial chromosome distribution confirming their potential role as scaffold proteins [186, 187]. However, although these extensive studies have detailed the structure and composition of isolated protein scaffolds, a number of glaring issues remain to be resolved. Firstly, the composition and structure of intact scaffolds is absolutely dependent on the isolation conditions utilised [182]. For example, isolation is dependent on protein cross-links stabilised by specific metal ions. Alternative extraction methods produce extended DNA structures with no visible scaffolds or core structure [182]. This

has led some researchers to suggest that the scaffold structure is an artefact resulting from the isolation techniques used [189]. Secondly, *in vivo* imaging has revealed that the positioning of the scaffold protein topoisomerase II is extremely dynamic within the nucleus [190]. In fact, the vast majority of topoisomerase II bound to mitotic chromosomes *in vivo* does not behave as an immobilised component of a static structural framework [190]. This calls into question the validity of using isolated scaffold preparations to characterise structural scaffold proteins.

Recent studies from Poirier and Marko have led to the complete re-assessment of the radial loop model [191]. Using micromechanical force measurements they analysed the elastic properties of individual chromosomes exposed to either micrococcal nuclease or short-recognition-sequence blunt-cutting restriction enzymes [191]. Importantly, isolated chromosomes were always maintained in an extracellular buffer that was shown to have no impact on chromosome morphology or elasticity when compared to chromosomes *in vivo* [191]. Treatment with micrococcal nuclease resulted in the rapid loss of chromosome elasticity and finally to complete dissolution of the chromosome [191]. Similar results were obtained when four base pair cutting restriction enzymes were used in place of micrococcal nuclease [191]. These results led the authors to conclude that the mechanical integrity of chromosomes is largely dependent on DNA and not on an internal proteinaceous core [191]. Instead, they propose that chromosomes consist of a cross-linked network or mesh of 30-nm fibres [191]. By comparing the effect of treating chromosomes with either 4 base pair cutters or 5 base pair cutters they were able to speculate that the average distance between cross-links could be as little as 10-20kb [191].

Although these findings do not give a detailed understanding of mitotic chromosome structure they do preclude the possibility of a stable structurally continuous proteinaceous core scaffold.

1.4 The Nuclear Matrix

Using high-salt extraction and DNase I treatment of nuclei Berezeny and Coffey identified a proteinaceous skeleton in the interphase nucleus that they termed the 'nuclear matrix' (although strictly speaking they were not the first to isolate this cellular fraction, they where the first to publicly speculate on its significance) [192]. Electron microscopy revealed that this matrix consisted of a dense meshwork of fibres connected to the nuclear lamina [192]. Furthermore, the removal of chromatin in these preparations had little impact on gross nuclear topology [192]. These findings led to the hypothesis that an underlying matrix maintains the nuclear architecture and serves as a platform for the regulation of nuclear events. Since these initial observations numerous proteins and ribonucleoproteins have been identified as potentially contributing to the nuclear matrix [193]. For instance, NMPdb, an online database for nuclear matrix proteins contains details of at least 398 different proteins [193]. However, the field of the nuclear matrix is not without controversy. Opponents of the nuclear matrix suggest that the insoluble networks recovered after high salt extraction of nuclei result from the precipitation of otherwise soluble nuclear proteins and that they do not prove the existence of a nuclear matrix [194-196]. Others are more positive suggesting that these artefacts indicate the presence of a pre-existing nuclear matrix, albeit in a different conformation *in vivo* than that seen in extracted nuclei [197].

One particularly contentious issue is the suggestion that ribonucleoproteins play a structural role in the nuclear matrix. This idea arose from a number of observations. First, when the nuclear matrix is prepared without digestion with RNase, RNA can be released from the structure only by sheer force or subsequent RNase digestion [198, 199]. Second, released RNA is recovered in the form of ribonucleoproteins, leading to the suggestion that ribonucleoproteins are a structural component of the nuclear matrix [200]. Third, nuclear matrix preparations without RNase digestion yield a fibrous network composed of ribonucleoproteins [198]. However, these findings must be viewed with a degree of scepticism as data from Tan et al. indicate that these observations may be an artefact resulting from the extraction procedure [194]. For example, they show that two different soluble tetrameric ribonulcleoproteins can be induced to form regular helical filaments ranging in length from 100nm to >10μm by RNase digestion or salt extraction (conditions similar to those used in matrix preparations) [194]. At low concentrations these fibres range from 7-10 nm in diameter, but at protein concentration of >0.1mg/ml the filaments rapidly aggregate to form thick filamentous networks that resemble nuclear matrix preparations [194]. Interestingly, this is not an isolated observation as the HIV Rev protein that binds to viral pre-mRNA transcripts in the nucleus undergoes spontaneous filament formation when released from its normal RNA binding partner [201]. These observations call into question the degree to which nuclear matrix filaments reflect pre-existing versus extraction-induced structures.

Despite these arguments there is clear evidence for the presence of polymer forming proteins such as lamins and actin within the nucleus [202]. Unfortunately, there is as yet little information regarding their polymerisation state within the nucleus *in vivo*.

1.4.1 The Nuclear Matrix *in Vivo*

To avoid the limitations and potential artefacts resulting from matrix preparations a number of investigators have turned to live cell imaging in the hope of catching a glimpse of the nuclear matrix *in vivo*.

In one such study the matrix associated protein toposisomerase II was fluorescently tagged with GFP (as mentioned previously) [190]. Live cell imaging revealed that topoisomerase II was highly mobile within the nucleus, with no detectable fraction being attached or immobilised on an underlying matrix [190]. These results are in strict contrast with the original finding that topoisomerase II is stably associated with nuclear matrix preparations, again calling into question the validity of these extraction methodologies [185, 186].

Using a similar approach Nalepa and Harper finally provide direct evidence for a nuclear matrix *in vivo* [203]. They show that in live cells GFP tagged Cdc14B (a phosphatase important in cell cycle regulation) decorates long filaments that begin at the nucleolar periphery and extend to the nuclear envelope, frequently making connections with the nuclear pore complexes [203]. Using the actin depolymerisation agent latrunculin they show that this filament network is actin-dependent [203]. Furthermore, using FRAP they show that while nucleoplasmic Cdc14B is highly mobile, filament-bound Cdc14B is tightly associated with the underlying structure [203]. Importantly, immunohistochemistry confirmed that endogenous Cdc14B displays the same localisation pattern as the GFP tagged protein [203]. Finally they also show that Cdc14B is critical for the maintenance of proper nuclear architecture [203].

These data provide the first *in vivo* evidence for the presence of a filamentous network of actin within the nucleus, and indicate a possible role for Cdc14B in regulating the structural properties of this matrix.

1.4.2 The Future of the Nuclear Matrix

Early studies indicated that the nucleus was filled with a dendritic proteinaceous meshwork that fills the entire nucleus. Yet such a structure is inconsistent with the limited size and topography of the interchromatin space in living cells (see below) [196]. In these models how do the voluminous chromosomes fit within this compact meshwork? Or do these models suggest that the matrix permeates throughout the chromosomal territories forming part of a chromosome scaffold? Doubts have also been raised about whether the extraction methods cause the aggregation of normally soluble proteins into matrix like structures [194-196].

These experimental and conceptual hurdles have led a number of people to doubt the presence of any stable matrix within the nucleus. Instead, macromolecular crowding has been suggested as the principle organising agent within the nucleus [204]. Yet although it is highly likely that macromolecular crowding has an essential role in the biology of the extremely concentrated nuclear milieu, it is hard to imagine that the co-ordination of nuclear events and the maintenance of nuclear architecture in the absence of some form of underlying matrix structure. It is also difficult to reconcile the directed and controlled movements of gene loci within the nucleus without a nuclear matrix from which these movements can be organised or launched (see below).

At last, *in vivo* live cell imaging has confirmed the presence of a matrix within the nucleus [203]. It is likely that further studies will reveal that the matrix *in vivo* bears little resemblance to early electron micrographs of isolated matrix preparations. Furthermore, many proteins identified as 'nuclear matrix proteins' are likely to be shown to have only transient interactions with the nuclear matrix, such as for topoisomerase II. Finally, further investigations will almost certainly reveal the matrix to be a highly dynamic entity that is in continuous flux. This possibility is perfectly reflected in the words of Jeffrey Nickerson on the subject of the nuclear matrix "Anything static is certainly dead" [205].

1.5 Scaffold/Matrix Attachment Regions

Establishment of the radial loop model led to the question of how the genome is ordered into loops and attached to the scaffold core. In metaphase nuclear scaffold preparations and interphase matrix preparations AT-rich DNA fragments are specifically retained [206]. It has been proposed that these sequences correspond to scaffold or matrix attachment regions (S/MARs) that bind the base of DNA loops to either the scaffold core or nuclear matrix, depending upon the stage of the cell cycle [207-209]. Consistent with this hypothesis, using AT-specific fluorescent stains, S/MARs can be visualised as long 'AT-queues' running the chromosome length [206]. In addition to their structural role in maintaining chromosome architecture, S/MARS are thought to be involved in a number of additional biological processes, such as regulation of gene expression/silencing, regulation of chromatin structure, and even chromatid separation (see below).

S/MARs are abundant in eukaryotic genomes, occurring every ~30-100kb, leading to the proposal that the genome is divided into loop domains of independently regulated chromatin [210]. The formation of these loops may form a physical barrier to prevent inappropriate interactions between proximal genes, and/or may have an insulating function to prevent the spread of heterochromatin between different 'chromatin domains'. This hypothesis is supported in part by a number of indirect observations. For example, S/MARS have been mapped to the borders of a 20kb DNase I hypersensitive region that includes the chicken lysozyme gene, suggesting that these elements demarcate the extent of open chromatin at this locus [211]. When these S/MAR sequences are included either side of a reporter gene in chicken cells they are able to protect the transgene from silencing imposed by sequences proximal to the integration site [211]. In the absence of these S/MAR sequences transgene expression is variable [211]. In addition, when placed between an enhancer and an adjacent promoter these elements are able to insulate the promoter from the enhancer [211]. Importantly, these properties are not unique to the chicken lysozyme S/MAR elements, as numerous other S/MAR sequences have been identified with similar functional properties.

Another example is the apolipoprotein B gene, which resides within a 47.5kb chromatin domain that is flanked by S/MAR elements [212]. The S/MAR elements demarcate the boundaries between open chromatin of the apolipoprotein B gene locus and closed chromatin outside of the locus [212]. Transgenes flanked by the 5' and 3' S/MAR elements are expressed in a position independent manner, but in the absence of the S/MAR elements transgene expression is variable and position dependent [212]. Importantly, the presence of S/MAR elements has no effect on the expression of transiently transfected (and therefore un-integrated) transgenes [212]. Together these

results suggest that the S/MAR elements act as domain boundary elements insulating the apolipoprotein B gene against position effects, and that S/MAR elements need to be integrated into the genome to exert their function [212].

However, independent experiments have revealed that the apolipoprotein B S/MAR elements may not function simply as boundary elements to demarcate domains of open/closed chromatin [210]. S/MAR sequences from the apolipoprotein B gene are able to protect mini-white transgenes from silencing in *Drosophila*, but only when two copies are included in the construct [210]. Intriguingly, insulation was still achieved when both S/MAR copies were included at the 5'end of the transgene [210]. Furthermore, deletion of the S/MAR sequences *in vivo* by site-specific recombination revealed that the S/MAR sequences were required for establishment but not maintenance of chromosomal insulation [210]. Together these results are not compatible with the chromosomal loop model in its simplest form suggesting an alternate mechanism of S/MAR function, at least in *Drosophila* [210].

It must be noted that not all S/MAR elements coincide with boundary/insulating activities. In *Drosophila*, two divergently transcribed hsp70 genes are contained within one functional unit that is demarcated by flanking 'specialised chromatin sequences' (scs and scs') sites [213]. These scs sites are able to insulate transgenes from silencing and have enhancer-blocking function [213]. In contrast, the region between the two hsp70 genes that contains a strong S/MAR site has no enhancer blocking function [213].

Perhaps an even more extreme example is the immunoglobulin intragenic μ enhancer, which is able to mediate transcriptional activation over large distances [214]. The

enhancer is flanked by S/MAR elements, yet interestingly, instead of blocking the activity of the enhancer these S/MAR elements are required for the enhancer to function in transgenic mice [214]. Furthermore, in transfection experiments these S/MAR elements can overcome methylation-induced repression of the long-range potential of the μ enhancer, and can even induce nucleosome acetylation at distal positions [214].

S/MAR elements can also have roles other than regulating gene expression. The Chinese hamster dihydrofolate reductase (*DHFR*) domain consists of two divergently transcribed genes (*DHFR* and *2BE2121*) that flank a 55kb intergenic spacer [215]. Replication initiates at any of a large number of sites scattered throughout the intergenic spacer, with three sites being preferred [215]. Interestingly, a S/MAR site is localised in close proximity to these preferred sites leading to the suggestion that attachment to the interphase matrix might be required for origin activity in early S phase [215]. This model would be consistent with studies showing that DNA synthesised at the onset of S phase remains close to the matrix throughout S phase, but DNA synthesised at later times migrates into the DNA loops [216, 217]. However, deletion of the *DHFR* S/MAR element has no effect on either the frequency or timing of the initiation of replication at this locus [215]. Rather, deletion of the S/MAR element interfered with the local separation of daughter chromatids, suggesting that there exists a sub-set of S/MAR elements that have important biological roles in regulating chromatid cohesion and separation [215].

In this respect S/MARs have been divided into two classes based on their biological activity [218]. Structural S/MARs are thought to maintain the chromosome structure and are permanently attached to the nuclear matrix [218]. Functional S/MARs often

contain tissue-specific sites for transcription factors and are differentially associated with the nuclear matrix, often in a tissue specific manner [218]. An example of a locus that includes both S/MAR sub-types is the murine *Igf2* locus, which contains a total of 4 S/MAR sites. S/MAR3, which lies 8kb downstream of *Igf2*, is constitutively associated with the nuclear matrix in all tissues [219]. In contrast, matrix association of S/MAR0 and S/MAR2 is tissue specific and developmentally regulated [219]. Interestingly, matrix association of S/MAR0 and S/MAR2 is restricted to the paternal allele, indicating that imprinting at the *Igf2* locus regulates these S/MAR elements [219].

1.5.1 Composition and Structure of S/MAR Elements

S/MARS are functionally conserved among eukaryotic species, and nuclear matrix proteins can bind S/MARs from different organisms [220]. Furthermore, S/MARs are functional *in vivo* when included in transgenes in other species. For example, as mentioned above human S/MAR sequences from the apolipoprotein B gene are able to suppress silencing in *Drosophila*, and the chicken lysozyme S/MARs are able to confer copy-number dependent expression in transgenic rice plants [210, 221]. However, there is very limited sequence conservation of S/MARs even within a species. It has therefore been suggested that S/MARs form specific secondary structures and that it is these secondary structures that are paramount in the recognition of S/MARs by S/MAR binding proteins and not the primary DNA sequence (for example see [222]). It is possible that these secondary structures are conserved across species, potentially explaining the observation that S/MARs can function in other systems in the absence of DNA sequence conservation [210]. S/MARs vary in size from several hundred base pairs to more than 10kb [210]. Although, there is no S/MAR consensus sequence, S/MARs share a number of different motifs, the most obvious feature is that S/MARs

contain ~70% ATs. S/MARs can also contain combinations of the following motifs [218, 223]:

- Kinked DNA generated by the presence of TG, CA, or TA dinucleotides separated by 2-4 or 9-12 nucleotides
- Intrinsically curved DNA
- Potential origins of replication
- Topoisomerase I and II binding and cleavage sites
- Homeotic protein-binding sites
- Polypurine and polypyrimidine stretches, which are known to form triple-helical or H-DNA structures
- Binding sites for transcription factors
- Sites of stress-induced DNA helix destabilisation
- MAR recognition signatures (MRS), which are bipartite sequence elements composed of two degenerate nucleotide motifs ‹200bp apart

1.5.2 S/MAR Binding Proteins

To define the mechanisms responsible for mediating S/MAR function it is important to determine which proteins are interacting with these sequences *in vivo*. This has led to the identification of a number of S/MAR binding proteins including lamins, ARDP, hnRNP-u/SafA, SatB1, Bright, and topoisomerase II [224-229].

As mentioned above, live cell imaging shows topoisomerase II to be freely mobile within the nucleus, suggesting that it is not an immobile, structural component of the

chromosomal scaffold or the interphase matrix, but rather a dynamic interaction partner of such structures [190].

SATB1 (special AT-rich binding protein 1) was originally identified as a protein that recognises double-stranded DNA with a high degree of base-unpairing (referred to as base-unpairing regions or BURs), which is a typical feature of S/MAR elements [226]. Fluorescent *in situ* hybridisation (FISH) analysis has revealed the presence of SATB1-bound genomic sequences (SBS) at the bases of DNA 'halos' generated by salt extraction of interphase nuclei, suggesting that SATB1 may have an important structural role in the interphase nucleus [230]. SATB1 is found predominantly in thymocytes and ablation by gene targeting results in arrested T cell development and death within 3 weeks of birth [231]. In *Satb1*-null thymocytes the temporal and spatial expression of hundreds of genes is dysregulated, including genes important for T cell function such as *Il2rα* and *Il7r* [231]. SATB1 targets multiple chromatin-remodelling complexes to specific genomic sites to regulate chromatin structure and gene expression [232]. For example, in the *Il2rα* locus SATB1 recruits the histone deacetylase contained in the NuRD chromatin-remodelling complex to an SBS and mediates specific deacetylation of histones within the locus [232].

In thymocyte nuclei SATB1 has a nuclear cage-like 'network' distribution that is resistant to salt extraction, indicating that SATB1 is a primary component of a subnuclear structure [233]. Furthermore, a series of randomly cloned SBSs all localise to the base of chromatin loops by tethering to this SATB1 network [233]. In contrast, in *Satb1*-null thymocytes, these sites appear within the loop regions and not attached to the matrix, indicating that the SATB1 network provides docking sites to actively tether

specific genomic sequences [233]. Importantly, tethering to the SATB1 network is essential for the correct regulation of genes that are located as far as 50kb from SBSs [233].

1.5.3 S/MARs and the Nuclear Matrix

Although the metaphase radial-loop configuration has recently been called into question, the model proposed by Poirier and Marko still includes a structural role for S/MARs as candidates for mediating cross-links between chromatin fibres. Conversely, although S/MARs have unquestionable physiological activity, direct evidence for a loop-and-scaffold-like arrangement during interphase is minimal. Furthermore, S/MAR sites can be electroeluted from matrix preparations generated in buffers that reproduce nuclear ionic and redox conditions, suggesting that they are not stably linked to a matrix *in vivo* [204].

Although it is clear that S/MAR sites have physiological activity, a number of principle issues need to be addressed. First, is the structural organisation of the genome into loops important/essential for the physical sub-division of the genome into independent regulatory domains? Second, what is the structure of the interphase nuclear matrix *in vivo* and how does this relate to the mitotic chromosome scaffold structure (i.e. are they the same structure in different conformational forms)? Third, are the S/MAR sequences that are isolated from or that bind to nuclear matrix preparations actually associated with the nuclear matrix *in vivo*? Fourth, do S/MARs link DNA directly to the nuclear matrix or is the DNA instead associated with transcription or replication complexes that are themselves bound to the matrix.

Finally, 'S/MARs' is a rather vague title for a family that encompasses such a functionally diverse range of regulatory elements. It is obvious that the family needs to be formally sub-divided based on biological function rather than grouped together under one title purely on their common ability to bind to matrix preparations or the presence of loosely shared sequence motifs.

1.6 Interphase Chromosomes and Nuclear Compartmentalisation.

On exiting metaphase, the discrete compact chromosomes decondense into their metaphase state. For a long time it was assumed that metaphase chromosomes exist as long thin tangled threads randomly arranged in the nucleus. However, over the last 20 years considerable data have revealed a highly organised and compartmentalised nucleus.

More than a century ago studies by Rabl and then Boveri suggested that chromatin was not randomly organised within the nucleus but occupied distinct territories [234, 235]. However, only recent advances in technology have allowed the confirmation that metaphase chromosomes are organised into discrete, non-overlapping 'territories' [236]. These chromosome territories are separated by a space called the interchromatin compartment, which contains the transcriptional machinery and splicing complexes [237].

Initial low-resolution morphological studies suggested that the chromosomal territories would be relatively impermeable to large complexes (such as the transcriptional machinery) leading to the proposal that actively expressed genes must be positioned on the surface of the chromosomal territories facing out into the interchromatin

compartment [238]. However, a number of independent findings have suggested that this model is too simple. First, measurement of protein motility in nuclei of living cells suggests that most proteins are highly mobile *in vivo* and that there are no significant zones of exclusion [239]. Second, a study in wheat nuclei has revealed that transcription (visualised by bromouridine-triphosphate incorporation [BrUTP]) occurs within the interior of chromosomal territories [240]. Additionally, immunolabelling of nascent BrUTP-labelled RNA together with chromosome painting in human fibroblasts has revealed the presence of RNA transcripts throughout chromosomal territories [241]. Third, high-resolution imaging of the chromosomal territories has revealed that they are permeated by a continuous network of channels and therefore resemble more a sponge like structure than a solid entity [242]. The chromosomal territories therefore have a significant internal surface area with the invaginations allowing the transcriptional machinery access to genes seemingly buried within chromosomal territories and also allowing nascent RNA transcripts to be rapidly exported from the nucleus once processed.

The notion that chromosome territories are completely isolated bodies with no significant intermingling is probably an over-simplification resulting from the limitations of current detection techniques. For example, although electron microscopy reveals that the borders of chromosomes are generally separated by the interchromosomal space, decondensed chromatin from neighbouring territories can be seen in contact in limited regions [242]. It is therefore likely that FISH techniques frequently used for chromosomal painting lack significant resolution potential to reveal the full extent of the chromosomal territories.

Indeed, certain active loci can reside extraterritorially on large (megabase) chromatin loops extending from the chromosome territory surface. For example, the major histocompatibility (MHC) region at 6p21 is able to extend out from the territory of chromosome 6 on large loops in response to activation of the MHC genes [243]. Similarly, the gene dense region containing the epidermal differentiation complex (EDC) on 1q21 forms extended loops in keratinocytes but not in lymphoblasts, in which the genes are silenced [244]. The biological significance of these loci adopting extended conformations situated away from the main body of the chromosome territories has yet to be determined. It is also not understood whether such looping is a requirement for transcription at these loci or the result of high levels of transcription. Finally, it will be interesting to determine where these chromosomal extensions are reaching out to.

1.6.1 Nuclear Organelles

In addition to chromatin, the nucleus contains a number of bodies or compartments that are often described as nuclear organelles. In contrast to cytoplasmic organelles, nuclear organelles are not enclosed within membranes but are formed by accumulations of specific proteins and ribonucleoprotein complexes [245]. Such accumulations may help generate high concentrations of associated factors thereby increasing enzymatic efficiency or complex assembly [245]. Furthermore, such nuclear compartmentalisation may act to physically separate or sequester mature/immature and active/inactive factors [245]. An increasing number of nuclear organelles have been identified and below is a brief description of some of the best characterised.

1.6.2 The Nucleolus

The nucleolus is the most prominent nuclear feature and can be easily visualised by bright-field microscopy. Nucleoli form around tandemly repeated clusters of ribosomal RNA (rRNA) genes. These loci are termed 'nucleolar organiser regions' (NORs) [246]. The nucleolus is the site of ribosomal RNA (rRNA) synthesis, processing, modification, and the assembly of the ribosomal subunits [246].

1.6.3 Splicing Speckles

Electron microscopy has revealed the presence of 'speckles' of variable size (0.1-1.8μM) within the interchromatin compartment [246]. These speckles have been termed 'splicing speckles' or 'Splicing factor compartments' because they are highly enriched in splicing factors [246]. However, they do not appear to be the primary sites of splicing, as they are not associated with transcriptionally active genes [247]. Instead, it is thought that the speckles serve as storage or assembly sites of the spliceosomal components [247]. The splicing speckles seem to reside within fixed sites of the nucleus [248]. However, this does not mean that they are inert bodies; indeed live cell imaging has revealed that they are highly dynamic compartments with splicing factors continuously being recruited from them to sites of active transcription [248].

1.6.4 Cajal bodies

Cajal bodies or coiled bodies are spherical structures ranging from 0.2-1μM that were originally termed 'nucleolar accessory bodies' because they are frequently associated with the periphery of the nucleolus [245]. Cajal bodies can move within the nucleus and are sometimes found in association with specific gene loci such as the histone genes [245]. The exact role of Cajal bodies has yet to be determined, although they are

believed to be important in the maturation of nuclear RNP complexes, including snRNPs and snoRNPs [249]. Newly assembled snRNPs and snoRNPs accumulate transiently within the Cajal bodies before moving to the nuclear speckles or nucleoli, respectively [249, 250].

1.6.5 Chromosomal Positioning

In certain cells chromosomes are positioned non-randomly within the nuclear space and occupy preferential positions relative to the centre of the nucleus and relative to each other [251].

It is not understood how chromosomal positions are maintained, however, one possibility is that the shear size of the chromosomes compared to the free space within the nucleus limits any gross random movements [251]. It is also likely that the chromosomes are physically tethered to immotile structures within the nucleus. One potential substrate for tethering is the nuclear envelope [251, 252]. Chromatin can interact either directly or indirectly within components of the lamina that line the nuclear envelope [253-255].

Recently it has been observed that the lamins are not restricted to the nuclear periphery. In fact, lamins have been found throughout the nuclear interior possibly forming the basis a nuclear matrix [197, 256]. It is therefore interesting to speculate that chromosomes may be attached, possibly via lamins, to an underlying framework or matrix that exists within the nucleus.

Chromosomes could also be restrained by tethering to subnuclear compartments [252]. For example, chromosomes bearing the tandemly repeated ribosomal genes cluster together to form the nucleolus [252]. Their stable association with the nucleolus prevents the movement of these chromosomes.

In addition it has recently been shown that pre-mRNA splicing factor compartments are positionally stable within the nucleus [248]. It is possible that these compartments form a steric obstacle to chromosomal movements within the nucleus [251].

What is the functional significance of chromosomal positioning? The example of the nucleolus offers one potential explanation. The nucleolus is formed by the positioning of specific chromosomes in a manner that allows the ribosomal genes from the different chromosomes to come together in one place within the nucleus [251]. This generates a specialised region within the nucleus containing concentrated levels of ribosomal processing and assembly factors, and therefore ensures highly efficient rRNA processing [251].

Although the nucleolus is an extreme example it indicates that chromosomal positioning may allow the generation of specialised nuclear 'neighbourhoods' thereby providing specific environments for the expression of co-regulated genes [251]. In contrast, association with heterochromatin is a powerful means of establishing gene silencing (see below). Chromosomal positioning may allow the formation of heterochromatin 'neighbourhoods', thereby facilitating specific gene silencing [251].

1.7 Dynamic Movements of Gene Loci within the Nucleus

Although chromosomes adopt relatively fixed positions within nuclei, they appear extremely fluid and motile structures within their respective territories [257]. In addition, gene activation and gene silencing events are accompanied by dynamic chromosomal movements that potentially allow/prevent access to the transcriptional machinery.

In the nucleus nascent RNA production and RNA Pol II molecules co-localise into discrete foci or 'transcription factories' [258, 259]. Analysis of different cell types has revealed that the number of these transcription factories is relatively low. For example, erythroid cells are calculated to contain 100-300 RNA Pol II foci [260]. In contrast, it is estimated that erythroid cells express approximately 4,000 genes [260]. Taken together these findings suggest that multiple genes may share individual transcription factories [260].

An elegant study from the lab of Peter Fraser indicated that specific genes from chromosome 7 are found to co-localise to the same transcriptional factories when they were actively transcribed [260]. However, when the genes were not undergoing active transcription they did not associate with the transcription factories and did not co-localise [260]. Based on their findings they suggest a model in which genes with transcriptional potential physically reposition within the nucleus to existing sites of ongoing transcription when activated, rather than assembling a new transcriptional site [260]. This concept questions the traditional view that active genes or promoters recruit RNA Pol II and instead suggests that active genes are physically recruited and translocate within the nucleus to pre-assembled RNA Pol II containing factories [260].

A subsequent study from the lab of Richard Flavell revealed that genes important for the divergent differentiation of TH1 and TH2 type cells from naïve $CD4^+$ cells, which lie on different chromosomes, specifically interact in a developmentally controlled manner [261]. In naïve $CD4^+$ T cells there is a strong interaction between the TH1 type *Ifnγ* gene on chromosome 10 and the TH2 type *Il5* gene on chromosome 11 [261]. When naïve $CD4^+$ T cells are forced to differentiate into effector TH1 and TH2 cells these interactions are diminished [261]. In contrast, non-T cell types do not exhibit these interactions. [261] It has been suggested that in naïve $CD4^+$ T cells, TH1 and TH2 type genes co-localise to a compartment within the nucleus that is permissive for expression, termed a 'poised chromatin hub', thereby priming both gene sets for rapid expression [261]. On differentiation into TH1 or TH2 type cells, genes specific for the opposing cell type are relocated away from the 'poised chromatin hub' and their expression prevented [261].

At the opposing end of the gene expression spectrum, it has previously been demonstrated that re-positioning of a gene to a region rich in centromeric heterochromatin is sufficient to induce gene silencing in *Drosophila*. The $brown^{Dominant}$ (bw^D) allele contains a null mutation resulting from the insertion of a block of heterochromatin into the coding region of the *brown* gene [262]. This results in the bw^D gene to be associated with centromeric heterochromatin and to be silenced [263]. What is more interesting is that in bw^+/bw^D larva, homologous chromosome paring results in the wild-type allele also being localised to the centromere [263]. This centromeric association causes a silencing of the normal *bw* gene, suggesting that the mere

localisation of the wild-type gene within close proximity of heterochromatin is sufficient to cause trans-inactivation [263].

Subsequently such phenomena have been noted to occur naturally in other systems. For example, in B lymphocytes the *Igh* and *Igk* immunoglobulin heavy and light chain genes are expressed in a monoallelic manner [264]. The expressed alleles occupy nuclear positions away from heterochromatin [265]. In contrast, within the same cells the silenced alleles are associated with heterochromatin [265].

Similarly in developing thymocytes the transcriptionally active *Rag-1* and *TdT* genes are positioned away from heterochromatin [266]. As the cells mature, transcription from the *Rag-1* and *TdT* genes is terminated and the genes are repositioned to pericentromeric locations [266]. Interestingly, the nuclear repositioning occurs after the termination of transcription, suggesting that this movement is the result rather than the cause of the gene-silencing event [266]. This repositioning is heritable and is stably transmitted through cell division. In the VL3-3M2 cells (a cell line derived from a thymic lymphoma) *Rag-1* and *TdT* gene silencing is not associated with repositioning to centromeric heterochromatin [266]. Importantly, silencing in these cells is transient and non-heritable highlighting the possibility that heterochromatin association may be responsible for maintenance of the silenced state [266].

It is currently unknown how these nuclear repositioning events are co-ordinated, however, data from the lab of Amanda Fisher suggest a potential mechanism [267]. FISH analysis has revealed that the transcription factor Ikaros is localised in discrete foci in the nuclei of B cells in close association with centromeric heterochromatin [267].

The nuclear proximity of 6 lymphoid-associated genes, λ5, CD2, CD4, CD8α, CD19 and CD45 to these chromatin-associated Ikaros complexes was analysed in B cell lines representing different stages of B cell development [267]. Remarkably, an inverse correlation between proximity to Ikaros complexes and gene expression was observed [267]. Together these data suggest that transcriptionally inactive genes are selectively targeted to centromeric heterochromatin possibly through association with Ikaros [267].

These findings led to the proposal of a model in which hypothetical 'recruiter' molecules (possibly Ikaros) bind in a sequence specific manner to recently silenced genes recruiting them to centromeric heterochromatin, thereby maintaining the silenced state [267].

Despite the correlation between recruitment of gene loci to heterochromatin and gene silencing, a number of studies suggest that repositioning to heterochromatin is not entirely sufficient to induce gene silencing. For example, the pre-B-cell specific λ5 gene is expressed, albeit in a heavily variegated manner, when targeted directly into pericentromeric heterochromatin in mice [268]. In both expressing and non-expressing clones the transgene remains closely associated with the periphery of centromeric heterochromatin, indicating that expression of the λ5 transgene does not require repositioning away from heterochromatin [268].

Importantly, just as gene expression is not always inhibited by juxtaposition to heterochromatin, silencing can occur without recruitment to heterochromatin. For example, in primary erythroblasts the human α-globin gene is active and positioned away from centromeric heterochromatin [269]. In human primary T lymphocytes and

immortalised B cells the α-globin gene is inactive but shows no association with centromeric heterochromatin [269]. Furthermore, even when integrated as a transgene into different chromosomal locations in mice the silenced α-globin transgene does not associate with regions of centromeric heterochromatin or non-centromeric heterochromatin [269].

Despite the correlation between recruitment of gene loci to heterochromatin and gene silencing, it is not clear in many cases whether the association with heterochromatin is the cause or the result of gene silencing. However, taking into account the available data it is obvious that the association of gene loci with heterochromatin has an important role in regulating gene expression. Furthermore, chromosome positioning plays an important part in this process by facilitating the formation of heterochromatin dense regions within the nucleus.

1.8 Chromatin, Position Effect Variegation and Locus Control Regions

1.8.1 Heterochromatin and Euchromatin

In 1928 Emil Heitz identified a type of chromatin that remains condensed throughout the cell cycle, which he termed heterochromatin [270]. In contrast, the rest of the genome undergoes rounds of condensation and decondensation at different stages of the cell cycle, and is termed euchromatin [270].

Other than remaining perpetually condensed throughout the cell cycle, heterochromatin has a number of additional characteristic properties. Ribonucleotide labelling experiments reveal that heterochromatin is transcriptionally inert (with a few exceptions) [271]. Heterochromatin is also late replicating relative to euchromatin [272, 273]. Interestingly, when origins of replication normally associated with heterochromatin are moved to euchromatin they become early replicating, suggesting that the timing of replication is determined by localisation to heterochromatin and not the origins embedded within it (reviewed in [274]). Similarly, early replicating origins become late replicating when moved to heterochromatin [275]. How heterochromatin is marked for late replication is not known, however, its relative inaccessibility may favour its late replication. Interestingly, the histone deacetylase HDAC2 only localizes to sites of late-replicating DNA, providing a potential mechanism to maintain heterochromatin structure through late replication [276].

Heterochromatin also carries a number of defining biochemical marks such as histone hypoacetylation, trimethylated H3-K9, monomethylated H3-K27, and in higher eukaryotes DNA methylation [277].

Besides these biochemical marks a number of proteins are associated with heterochromatin. The most characterised is heterochromatin protein-1 (HP1), which is found in various organisms ranging from *S. pombe* to mammals [278]. In mammals there are three HP1 isoforms, HP1α, HP1β and HP1γ which although share similarities in amino-acid sequence and structural organisation display differences in nuclear localisation [278]. For example, HP1β and HP1γ co-localise to regions of heterochromatin and euchromatin, whereas HP1α is almost exclusively localised to heterochromatin [279, 280]. HP1α binds to trimethylated H3-K9 through its chromodomain and facilitates the formation of heterochromatin [46, 47]. Interestingly, treatment of mouse cells with RNase disrupts the localisation of HP1α, and simply adding back purified total of nuclear RNA (but not *E.coli* RNA) restores HP1α localisation, indicating that specific RNAs may play a structural role in heterochromatin formation mediated by HP1 [281].

The increased cytological visibility of heterochromatin over euchromatin suggests that heterochromatin forms a far more condensed structure than euchromatin. This prediction is supported by the findings that heterochromatin is far less accessible to digestion by nucleases and sediments faster than euchromatin through a sucrose gradient [282]. The compact nature of heterochromatin is further exemplified by the finding that an antibody raised against a 'branched chain' of four identical linked methylated H3-K9 peptides is specific for pericentric heterochromatin [281]. This result

is significant as only the close association of dimethylated H3-K9 residues on independent nucleosomes (i.e. within extremely condensed chromatin) could form the epitope for the 'branched chain' type antibody [281]. In contrast, an antibody raised against a linear peptide decorates chromosomes in a uniform manner [281].

Electron microscopy has revealed that the increased density of heterochromatin may form from the lateral association of chromatin fibres [283]. Heterochromatin isolated from granulocyte nuclei features thick 45-50nm electron dense structures apparently origination from the folding of 30nm fibres back on themselves [283]. The diameter of this folded back structure is less than what would be expected for two 30nm fibres lying side by side, indicating that the folding involves mutual intercalation of nucleosomes between laterally interacting fibres [283]. This extensive lateral association and intercalation of heterochromatin fibres has been observed in numerous species [283]. Interestingly, the regular spacing of nucleosomes by certain chromatin remodelling complexes has been associated with gene silencing [284]. Such regular spacing would facilitate intercalation and lateral association of chromatin fibres, thus providing a potential mechanistic link between ATP-dependent chromatin remodelling and chromatin condensation.

A number of additional proteins have been identified with the potential to mediate lateral association of chromatin fibres. For example, recent studies have shown that MeCP2, a protein that binds methylated DNA and heterochromatin, is able to cause extensive compaction and self-association of defined nucleosomal arrays *in vitro* [285]. There is also evidence that HP1 can form internucleosome dimers in *trans*, thereby potentially linking two separate chromosome fibres [286]. The recent finding that HP1

may bind to two H3-K9 residues may also be relevant to the cross-linking of chromatin fibres [287]. Polycomb group (PcG) repressor proteins form multimeric complexes that are required for the developmentally regulated silencing of genes that determine cellular fate. Evidence suggests that PcG proteins may function in a similar way to HP1 in promoting lateral fibre association [288]. Finally, the repressor protein MENT (myeloid and erythroid nuclear termination stage-specific protein) has been shown to drive nucleosome array bridging *in vitro* [289].

1.8.2 Constitutive and Facultative Heterochromatin

Heterochromatin can be divided into two types, constitutive heterochromatin and facultative heterochromatin [290].

Constitutive heterochromatin is formed early in embryogenesis before the beginning of zygotic transcription and remains inactive throughout the rest of development [290]. Constitutive heterochromatin occurs primarily in large blocks near centromeres and telomeres and consists mainly of repetitive satellite DNA, which is composed of short sequence units that are tandemly repeated up to millions of times [290]. In addition, constitutive heterochromatin contains a considerable quantity of transposable element-related sequences. It is thought that the heterochromatinisation of these sequences acts to prevent the illegitimate recombination between repeats that would otherwise lead to deleterious chromosomal rearrangements, and may also neutralise transposable elements embedded within them [169]. This is supported by the finding that mutations affecting the regulation of chromatin structure result in the mobilisation of transposable elements in *Drosophila* and *Arabidopsis* [291-294].

Constitutive heterochromatin forms approximately 20% of the genome in humans, 30% in *Drosophila melanogaster* and up to 85% in certain nematodes [290]. Interestingly, almost all heterochromatin can be specifically eliminated from somatic nuclei during early development of nematodes and other organisms [295]. Furthermore, in certain polytene tissues of *Drosophila* much of the constitutive heterochromatin fails to undergo DNA replication, and therefore becomes dramatically underrepresented [290]. These findings have led to the suggestion that heterochromatin is a genomic wasteland consisting of junk DNA with little or no functional significance. However, this idea is slowly being replaced by the concept that constitutive heterochromatin is an important genomic compartment that plays a structural role in chromosomal and nuclear architecture, as well as providing a mechanism for regulating gene expression [290]. For example, in *Drosophila melanogaster* constitutive heterochromatin is essential for centromeric activity/stability and chromosomal pairing [296-298]. Intriguingly, the extremely repetitive DNA that forms the basis of constitutive heterochromatin is highly divergent between species and lacks coding sequences [299]. This suggests that the primary DNA sequence may be unimportant for function and that blocks of satellite DNA do not need to form specific structures, but instead act as passive spacers that define the centromere [300]. In this case, the only requirement could be the formation of heterochromatin domains around the centromere, a role that could potentially be fulfilled by any repetitive sequence.

Furthermore, although constitutive heterochromatin can be eliminated from certain somatic cells, the fact that it is always preserved in germline cells indicates some physiological advantage from its retention [290].

Finally, a number of the expressed genes have been discovered embedded within domains of constitutive heterochromatin [301-304]. Intriguingly, these genes fail to be expressed correctly when moved to euchromatin by chromosomal rearrangements, indicating that they depend on a heterochromatic location for correct expression [305, 306]. How expression of these genes is achieved from within domains of heterochromatin is not known. However, this does call into question the assumption that constitutive heterochromatin is a homogeneously condensed, inaccessible, transcriptionally inert structure. Because these genes are not expressed when present in euchromatin it is tempting to speculate that proteins, which are essential for their correct expression, are specifically enriched in heterochromatin. This hypothesis seems plausible for reasons that will be briefly discussed below.

1.8.3 Facultative Heterochromatin

In contrast to constitutive heterochromatin, facultative heterochromatin corresponds to the developmentally regulated, epigenetically stable, silencing of potentially euchromatic sequences [290]. During development, genes that no longer need to be expressed often become condensed to form facultative heterochromatin, most likely as a mechanism to prevent their unsolicited reactivation. As a result, the proportion of the genome packaged into facultative heterochromatin gradually increases as development proceeds, until in terminally differentiated cells such as granulocytes most of the genome is condensed [283].

The best-characterised example of facultative heterochromatin is the inactive X chromosome in somatic cells of female mammals (reviewed in [307]). Here the double complement of X chromosomes is compensated by the random silencing of one X

chromosome [307]. Once silenced the X chromosome remains stably condensed throughout subsequent cell division [307]. X inactivation is initiated by the expression of an untranslated RNA called Xist (X-inactivation-specific transcript) from the X-inactivation centre (Xic) of the chromosome that will be silenced [308]. The Xist RNA coats the chromosome in *cis* and drives inactivation and heterochromatinisation of the whole X chromosome [307]. Inactivation is accompanied by the hypoacetylation histone H3 and H4, dimethylation of H3-K9, and DNA methylation at the 5' ends of genes. In addition, the inactive X chromosome becomes late replicating and enriched in the histone H2A variant macro-H2A [309, 310]. Thus, facultative heterochromatin carries similar biochemical marks to constitutive heterochromatin with only few subtle differences.

Interestingly, the diverse sequence composition of facultative heterochromatin highlights that specific sequences are not absolutely required for the formation of heterochromatin, although repetitive sequences may play an important role in initiating heterochromatin formation (see below).

1.9 Heterochromatin Formation and RNAi

RNA interference (RNAi) is an evolutionary conserved silencing mechanism that is triggered by double stranded RNA (dsRNA) and serves to regulate gene expression at many levels, including targeted degradation of mRNAs, repression of mRNA translation and direct repression of transcription [299, 311, 312]. Silencing is initiated through the formation of double stranded RNAs (dsRNA) generated from repetitive sequences, either as a result of 'sense' and 'anti-sense' transcription of the same template, or in some organisms (fungi, nematodes and plants) as a consequences of

RNA-dependent RNA polymerase (RdRP) activity [313]. Respective dsRNAs are cleaved by the RNase-III-like enzyme Dicer generating small interfering RNA (siRNA) molecules ~22 nucleotides long [314]. The siRNAs generated by Dicer associate with a multiprotein RNAi effector complex called RISC (RNA-induced silencing complex) to provide specificity for degradation of complementary mRNAs [315].

Recent findings have also connected the RNAi silencing pathway directly to the formation of heterochromatin. For example in *S. pombe* deletion of components of the RNAi machinery (namely Dicer, RdRp, and Argonaute [named *dcr1*, *rdp1* and *ago1* in *S. pombe*]) leads to the disruption of H3-K9 methylation, decreased binding of Swi6 (an HP1 homologue) to heterochromatin and the loss of heterochromatin induced silencing [316]. Similarly, in *Drosophila* deletion of components of the RNAi machinery results in defects in heterochromatin assembly, correlating with loss of histone H3-K9 methylation and disruption of HP1 localisation [292]. Likewise, loss of dicer function in a chicken-human hybrid DT40 cell results in cell death, premature sister-chromatid separation and abnormalities in the localisation of heterochromatin-associated proteins [317]. Finally, in *Arabidopsis* both DNA methylation and H3-K9 methylation are dependent on the RNAi machinery [176, 177]. Thus, it appears that the involvement of RNAi in chromatin regulation is an evolutionary conserved mechanism in a number of species.

An important feature of RNAi-mediated silencing is the generation of siRNAs homologous to the target loci [311]. Recent studies in wild type *S. pombe* have identified small RNAs, resembling Dicer cleavage products, which show homology to centromeric repeats [318]. It has been proposed that centromeric repeats are transcribed

in both directions to generate overlapping non-coding dsRNAs that are then processed by the RNAi machinery [316]. The resulting siRNAs then promote the methylation of H3-K9 at the corresponding centromeric regions, followed by Swi6 recruitment and heterochromatinisation [316]. Interestingly, *Dicer* mutations in both *S. pombe* and chicken-human hybrid DT40 cells result in the accumulation of overlapping centromeric transcripts, providing evidence that centromeric repeats are transcribed at low levels [316, 317]. This finding is clearly at odds with the classical view that heterochromatic repeats are transcriptionally inert. With this in mind it is also interesting to note that in *S. pombe* Rdp1 (an RdRP) is physically associated with centromeric heterochromatin, ideally located to ensure the continued regeneration of dsRNA and siRNAs to maintain silent heterochromatin [319, 320].

Remarkably, further studies in *S. pombe* have shown that expression of synthetic hairpin RNAs is sufficient to silence corresponding homologous loci in *trans*, and to induce the assembly of silent heterochromatin displaying all the characteristic modifications such as methylation of H3-K9 and binding of Swi6 [321]. Importantly, this silencing is dependent upon both the RNAi machinery and Clr4 (an H3-K9 HMT) [321].

How siRNAs direct recognition and silencing of the appropriate region remains unclear, but pairing between siRNAs and either homologous DNA sequences or nascent RNA transcripts at the target locus provide a potential mechanism to impart specificity.

Interestingly, RNAi is not universal, as *S. cerevisiae* lacks components of the RNAi system as well as Swi6/HP1 and specific H3-K9 HTMs [322]. Interestingly *S. cerevisiae* has a far lower repeat content than *S. pombe* and also lacks heterochromatic

centromeric repeats. This suggests that the RNAi-induced heterochromatinisation may have evolved as a 'surveillance' mechanism to neutralise repetitive or parasitic DNA sequences [323]. It also interesting to note that in *S. pombe* it appears that the RNAi-mediated repeat silencing has been hijacked to allow regulation of endogenous repeat proximal genes that are involved in meiosis [321].

1.10 Position Effect Variegation in *Drosophila*

In 1930, shortly after the characterisation of heterochromatin by Heitz, Hanz Muller identified the gene silencing potential of heterochromatin in *Drosophila* [324]. Genetic analysis revealed that an X-ray-induced chromosomal rearrangement had brought the white gene, responsible for the red eye pigmentation, into proximity with pericentric heterochromatin resulting in failure to express the white gene in clusters of cells. It was proposed that this proximity to heterochromatin induced the silencing and heterochromatinisation of the white gene in a proportion of the cells resulting in mosaic pigmentation of the eye [325]. The mosaic clusters of pigmented and unpigmented cells displayed by the mutants implies that a decision to activate/silence the white gene is taken relatively early in eye development and then this decision is stably maintained, possibly by epigenetic modifications, through successive cell divisions [3, 326, 327].

Subsequent studies have revealed that transgenes are also sensitive to heterochromatin induced silencing in *Drosophila* [328-330]. In this case, integration of the transgene proximal to regions of heterochromatin results in transgene silencing in a proportion of cells.

This phenomenon now termed position effect variegation (PEV) can be generally described as the 'stochastic, epigenetically stable, silencing of a gene in a proportion of cells that would normally be expected to express it' [331]. On silencing, the previously euchromatic DNA is forced to adopt the characteristics of heterochromatin, i.e. decreased accessibility to DNase I or micrococcal nuclease, late replication, and heterochromatin associated epigenetic modifications [178, 275, 332-334].

Although the exact mechanisms driving PEV are not entirely understood it has been suggested that (in a percentage of cells) heterochromatin proximal to the rearrangement site (or integration site for transgenes) can 'spread' across the breakpoint into the translocated gene segment thereby silencing the gene [3]. The finding that genes more proximal to heterochromatin are silenced prior to genes at more distal locations supports the spreading hypothesis [277]. In this case variability in the heterochromatin spreading distance would give rise to the variegated phenotype. A potential mechanism for spreading is revealed by the finding that HP1 can bind to both SU(VAR)3-9 (an H3-K9 specific HMT) and methylated H3-K9, suggesting that heterochromatin bound HP1 could recruit SU(VAR)3-9 to methylate adjacent nucleosomes, in turn resulting in the recruitment of additional HP1 and subsequent HMT activity [335].

Although the cis-spreading model of heterochromatin-induced gene inactivation is certainly correct in many circumstances, it cannot easily explain all features of PEV [3]. For example, some chromosomal rearrangements display heterochromatin-induced variegation that occurs over distances of several megabases [3]. Such long range influences are difficult to explain simply by the cis-spreading of heterochromatin along the chromatin fibre in a linear manner [3]. Furthermore, as mentioned above,

heterochromatin is able to induce trans-inactivation of the bw gene in *Drosophila* [263]. It has therefore been proposed that different regions of heterochromatin can interact in *trans*, and that these interactions may regulate the spatial organisation of chromosomes within the nucleus in a manner which could influence PEV [263, 330, 336]. Silencing mediated in this manner seems probable as heterochromatin forms a sub-nuclear compartment within the nuclei. Recruitment to this sub-nuclear compartment (possibly through homologous pairing) could drive heterochromatin formation in *trans* [263, 330, 336]. In this case PEV would result from the variable ability of a gene to be recruited to either heterochromatin or euchromatin compartments.

1.10.1 Modifiers of PEV

PEV is exquisitely sensitive to the concentration of proteins involved in the regulation and maintenance of chromatin structure [2]. An increase in the nuclear concentration or activity of proteins that favour the formation of heterochromatin augments PEV induced silencing (see below). This provides supporting evidence for the hypothesis that recruitment of genes to heterochromatin sub-nuclear compartments (that are enriched in heterochromatin proteins) is sufficient to drive PEV induced silencing. In contrast, an increase in the nuclear concentration of proteins that favour the formation of euchromatin will lessen the silencing effect of PEV.

Genetic screens in *Drosophila* have exploited the sensitivity of PEV to concentration of 'chromatin regulators', resulting in the identification of a large number of proteins that act as dose-dependant modifiers of variegation [2, 337]. These modifiers can be divided into two classes based on how they affect the variegation phenotype, suppressors of variegation [Su(var)s] and enhancers of variegation [E(var)s] (it is important to

remember that *Drosophila* gene names often reflect the phenotype generated by their deletion or inactivation) [2]. An example of a suppressor of variegation is the SU(VAR)3-9 protein, which is able to methylate H3-K9 [335]. Decreasing the dose of this protein reduces PEV, conversely, increasing the dose increases PEV [338]. Alterations in *Su(var)2-5*, which encodes for the protein HP1, affects variegation in a similar dose dependant manner [339].

Interestingly, modifiers of PEV have the opposite effect on genes originating from heterochromatic locations, potentially supporting the hypothesis that correct expression of these genes is dependant on proteins that are enriched in heterochromatin [48].

1.11 Position Effect Variegation in Mammals

More recently the advent of transgenic technology has led to the identification of PEV in mammalian systems. Transgenic studies have indicated that inclusion of enhancer and promoter elements is often insufficient to ensure appropriate transgene expression [340]. Frequently transgenes are either not expressed at all or are expressed at low levels and only in a proportion of cells that express the endogenous gene. In addition, transgenic lines harbouring the same transgene integrated at different sites within the genome can display a diverse range of phenotypes. It is now generally accepted that, when a gene is removed from its endogenous genomic location and is inserted at an ectopic site as a transgene, it becomes sensitive to both positive and negative influences at the site of integration. Importantly some integration sites are more likely to perturb transgene expression than others [341]. For example, transgene integration into regions rich in heterochromatin, such as centromere or telomeres, frequently results in PEV or complete transgene silencing. Mammalian PEV is likely to be mediated by similar

mechanisms responsible for regulating *Drosophila* PEV [331, 342]. Indeed, homologues of *Drosophila* proteins (that were originally characterised due to their ability to modulate PEV) have been identified in mammals [343-347]. Moreover, these proteins appear to be functionally conserved and are able to regulate PEV in mammalian systems [342].

1.12 Locus Control Regions

Transgenic technology has also led to the identification of a class of regulatory elements that are able to protect against PEV [348]. Locus control regions (LCRs) are powerful cis-regulatory elements that are functionally defined through transgenesis by their ability to render linked transgenes insensitive to PEV. LCRs are able to direct tissue-specific, copy number-dependent transgene expression in a manner that is independent of the site of integration [348]. Although functionally LCRs are a relatively novel concept, structurally they incorporate properties of other gene regulatory elements including promoters, classical enhancers, chromosomal insulators and matrix attachment regions [349]. Thus, despite the fact that more than 32 LCRs have been identified, sequence analysis has failed to reveal the presence of any LCR specific consensus sequences [349]. However, one feature common to all LCRs (although not restricted to LCRs) is the presence of tissue-specific DNase I hypersensitive sites (HSSs) which are thought to represent binding sites for regulatory proteins [349-351].

1.12.1 The β-globin LCR

The human β-globin gene-complex consists of 5 β-globin genes (5'-ε-Gγ-Aγ-δ-β-3') that are oriented in order of their expression in erythroid development [352]. Expression of the β-globin genes is regulated by an LCR that lies upstream of the globin gene

complex. The β-globin LCR was the first to be characterised and its identification resulted from a culmination of experimental observations. First, analysis of Dutch thalassaemia, which is caused by a deletion of 100kb (including the ε-, γ- and δ-genes but leaving the adult β-globin gene intact) and results in failure to express the adult globin genes, indicated that an important regulatory element exists within the deleted region [353, 354]. The Dutch deletion results in a closed chromatin configuration that engulfs the remaining globin gene complex suggesting that the deleted sequences normally function by forming a domain of open chromatin [353]. Subsequently, analysis of Hispanic thalassaemia, which is caused by a smaller deletion of ~35kb of sequences upstream of the globin gene complex, that also results in the formation of a closed chromatin conformation spanning the entire locus and the failure to express any of the globin genes, indicated that this proposed chromatin opening elements reside within a 35kb deleted region immediately upstream of the globin gene complex [355]. Second, early transgenic mice containing only the β-globin gene and proximal sequences expressed transgenes at levels significantly lower than that of the endogenous globin genes [356-358]. Furthermore, transgene expression was only detected in a small proportion of transgenic lines. Thus, as the transgenes were so sensitive to PEV this supported the notion that additional regulatory elements might exist in the native locus. Third, five developmentally stable DNase I HSS were identified immediately upstream of the globin gene complex approximately 6-22kb upstream of the ε-globin gene [359, 360]. Moreover, four of these HSS were found to be erythroid specific, signifying the presence of a tissue specific regulatory element within this region [360]. Finally, inclusion of these sequences in β-globin transgenes resulted in high levels of tissue specific expression in a position-independent and copy number dependent manner, a characteristic indicative of the presence of an LCR [348]. Thus, data from both the

Hispanic deletion and transgenic studies suggests that the β-globin LCR acts as a dominant chromatin-opening element within the globin gene complex. In contrast, a recent publication reports that a 150kb YAC spanning the entire β-globin locus is sensitive to PEV, as revealed by flow cytometry, when single copies were integrated at a number of different positions [361]. This last result indicates that the β-globin LCR may not be as dominant over PEV as once thought.

Additional transgenic studies have revealed that the activity of the β-globin LCR is tissue specific. For example, the expression of both globin transgenes and non-globin transgenes is confined to the erythroid cells when linked to the β-globin LCR [348, 362, 363]. Interestingly, however, in some cases ectopic expression has been observed in transgenic mice [363]. In these instances, although the LCR conferred high levels of position independent, erythroid-specific expression, the natural function of the linked non-globin promoter allowed variegated expression outside of the erythroid compartment [363]. Therefore, tissue-specific control of basal transcription appears to reside in the promoter element, whereas tissue specific enhancement resides in the LCR.

Interestingly, the sequence, arrangement of gene and regulatory elements and chromatin structure of the β-globin locus are highly conserved among mammals [364]. Conservation is present within the LCR, with the highest homology residing within the HSS forming sequences [364]. A total of six conserved HSSs have now been described in mouse and human, and homologues to at least 4 have been identified in other mammals [365, 366]. Sequence and structural homology between species extends upstream of the LCR and downstream of the coding regions, including several orthologous olfactory receptor genes that flank the human and murine globin gene

complexes [364]. A pair of HSSs (5'HSS-60/-62) upstream of the LCR embedded within the olfactory gene complex, and a single HSS (3'HSS1) downstream of the globin gene complex are also conserved between mouse and human [367-369]. Such a high degree of homology between species suggests that the LCR serves an absolutely essential role.

In vitro assays reveal that each HSS displays unique properties. For example, 5'HS2 functions as a classical enhancer in transient transfection assays, whereas the transcriptional enhancing activities of 5'HS3 or 4 only become apparent on integration into chromatin [349]. In addition, 5'HS2 is able to position a linked transgene away from centromeric heterochromatin to a compartment in which transcription is favoured [370]. The 5'HS5 functions as a chromatin insulator and enhancer-blocking element in both the human and murine LCRs, a role that has been shown to be CTCF dependent [368]. The role of 5'HS1 has not yet been determined. Interestingly, transgenic experiments indicate that all HSSs are required for appropriate transgene expression, and that the HSSs function in a cooperative manner [371].

1.12.2 *In Vivo* LCR Deletions

The *in vivo* deletion of individual HSSs from the murine globin LCR has yielded different results than the earlier transgenic studies [365, 372-378]. For example, analysis of expression levels in these mutants reveals that the HSSs function in an additive rather than cooperative manner and that, in contrast to transgenic studies, no HSS is absolutely required for a specific function [375]. This indicates that there is potentially a large degree of redundancy within the LCR when situated at its endogenous location.

In addition, a number of results have indicated that 'chromatin-opening' at both the endogenous mouse and human globin gene complexes occurs independently of the LCR, which is in stark contrast to data obtained from transgenic experiments. For example, Epner et al. deleted the murine LCR in ES cells, which were then forced to either differentiate in culture, or the mutant locus was transferred into human erythroid cells [373]. Analysis in both cases revealed that the general DNase I sensitivity of the globin locus was unaltered by the deletion, although expression of the globin genes was greatly decreased (5-25% of normal) [373]. In a similar study Schübeler et al. deleted the β-globin directly from the human chromosome 11 which had been artificially transferred into murine ES cells [379]. Following subsequent transfer of the mutant chromosome into mouse erythroleukaemia (MEL) cells the globin locus was seen to adopt an open chromatin configuration, as defined by DNase I hypersensitivity, in the absence of the LCR [379]. Furthermore, localisation away from centromeric heterochromatin was maintained in the absence of the LCR [379]. Together, these two independent findings have been interpreted to indicate that chromatin opening of the globin locus is not entirely dependant upon the LCR.

Importantly, these two '*in vivo*' deletion experiments share one potential caveat. That is, in neither case does the mutant locus undergo the epigenetic modifications normally imposed on germline transmission. Thus, by avoiding this developmental stage the physiological function of the LCR may have been bypassed. The observation that Pol II is already found in association with the globin LCR in undifferentiated ES cells validates this criticism [380]. Importantly, on transfer of a human chromosome 11 containing the Hispanic deletion into the MEL cells the globin locus adopted a closed

chromatin conformation and associated with centromeric heterochromatin as expected, confirming that the discrepancies are not the direct result of the extensive chromosome manipulations [379].

Bender et al. used a more physiological approach by deleting the whole mouse globin LCR in ES cells before generating mice [381]. Here the mutant locus passes through the germline and as a result is subjected to the epigenetic modifications. In these mice the globin genes are expressed in a developmentally correct manner, albeit at extremely low levels [381]. In addition, the mutant locus was shown to reside in a DNase I-sensitive conformation [381]. These results indicate that the LCR, as currently defined, is not required for the initiation or maintenance of an open chromatin structure or for the developmentally regulated expression of the native locus during somatic development and differentiation [381].

How can the closed chromatin conformation observed in Hispanic deletion patients, and the evolutionary conserved 'chromatin opening' property of the globin LCRs (in transgenic studies) be reconciled with the contrasting experimental results from the experimental deletion experiments? It is possible that there is significant redundancy within the globin complex and that sequences remaining within the mutant locus are sufficient for mediating chromatin opening. This explanation seems plausible, as the *in vivo* LCR deletion experiments do not remove all of the homologous sequences residing within the Hispanic deletion. Furthermore, the LCR has been defined as the minimum sequence preventing PEV in transgenic studies. Thus, the true extent of the LCR at the endogenous locus has not been identified, it therefore seems reasonable to suggest that the sequences defined as the LCR need to be expanded to include additional upstream

elements. In addition, all of the above observations were made on bulk populations of cells, and as a result variegation was never addressed (it must be noted, however, that more recent data using single cell RT-PCR provides evidence, although limited, that there is no variegation in these mice [382]). In such a case low levels of variegation would be seen as a mild decrease in the total level of expression (as observed in the small HSS deletions). Furthermore, opening chromatin and transcriptional enhancement may be separable functions of the LCR. In this manner a proportion (although not all) of the cells may be able to escape silencing and open the globin locus in the absence of the LCR, thus accounting for the retention of general DNase I sensitivity of the bulk population. However, even though the globin locus is open in a majority of cells the level of expression is extremely low due to the loss of enhancer elements from within the LCR.

1.12.3 Mechanisms of LCR Function

How the β-globin LCR regulates expression of the globin genes is not known in detail, however, two main models have been proposed, neither of which can explain all of the available experimental data [383]. The 'looping' model suggest that the 5' HSS's of the β-globin LCR fold to form a holocomplex, with the HSS core elements forming an active site that binds transcription factors. This structure physically loops out so that the LCR holocomplex comes into close proximity to the appropriate promoter [350]. Close association of the holocomplex with the promoter sequences allow the delivery of LCR-bound transcriptional proteins to the basal transcriptional machinery already bound at the promoter thus enhancing transcription.

Compelling support for the holocomplex model is provided by two different experimental approaches, RNA TRAP and chromosome conformation capture (3C) technology, which give tangible evidence that the LCR is in close association with the globin genes in expressing tissues [384-386]. The structure formed by these intimately associated regions has been termed an active chromatin hub (ACH) [387]. A recent study by Palstra et al. shows that in mice and humans a core ACH is developmentally conserved and consists of the HSSs 1-6 of the LCR, the upstream 5'HSS-60/62, and downstream 3'HSS1 [385]. Furthermore, in mouse erythroid progenitors that are committed to but do not yet express β-globin, only the interactions between 5'HSS-60/-62, 3'HSS1 and the HSSs at the 5' of the LCR are stably present [385]. After induction of differentiation, these sites physically cluster with the rest of the LCR HSSs and the globin genes, concomitant with globin transcription [385]. Interestingly, indirect evidence suggests that holocomplex formation may be directed by S/MAR sequences within the globin gene complex and LCR [388].

Further support for the looping model comes from *in vitro* experiments, which have shown that Pol II can be transferred in *trans* from immobilised LCR templates to a β-globin gene in a process that is facilitated by the erythroid transcription factor NF-E2 [380].

In contrast, the 'tracking' model suggests that transcription factors and transcriptional co-activators bind to the LCR, forming an activation complex that tracks linearly along the DNA helix of the locus. When this transcription complex encounters the basal transcriptional machinery located at the correct promoter transcription is initiated [350].

Support for a 'tracking' based model comes from recent data which suggest that intergenic transcription may play an important role in regulating globin gene expression [389-391]. In a subset of both murine and human erythroid cells the β-globin LCR and intergenic regions are unidirectionally transcribed by RNA Pol II [389]. Furthermore, the globin gene complex is divided into three differentially activated and developmentally regulated sub-domains, each of which is precisely delineated by intergenic transcripts [390]. Mutations that delete the start site of the adult-specific intergenic transcripts lead to a decrease in β-globin gene transcription and a reduction in general DNase I sensitivity, suggesting that intergenic transcription is responsible for modulating the chromatin structure of globin locus, preparing each sub-domain for appropriate stage specific gene expression [390].

Thus, in conclusion it appears that both looping and tracking mechanisms may function to regulate expression of the globin genes.

1.12.4 The hCD2 LCR

The second LCR to be defined is responsible for regulating expression of the human CD2 (hCD2) gene, which in humans encodes for a cell surface glycoprotein that is expressed on thymocytes, T cells and a small percentage of natural killer cells [392]. The 'accessibility' of the immune system and the availability of hCD2 antibodies allow the analysis of gene expression at the single cell level, which, in combination with the relative simplicity of the hCD2 locus, provides an ideal model to study LCR directed gene expression.

Originally a 29.4kb genomic fragment containing the five exons of hCD2 as well as 5.3kb upstream and 9.3kb downstream flanking sequences was introduced into the genome of transgenic mice (now called the hCD2 maxigene, see figure 1A). Analysis of both RNA and protein levels revealed that transgene expression was tissue specific, copy number dependent and independent of the site of integration [392]. This finding indicated that all the regulatory elements required for correct expression of hCD2 were present in the 29.4kb fragment and that these elements function as an LCR [392]. To identify the minimal sequences required for appropriate hCD2 expression a hCD2 construct was generated consisting of 5.3kb of upstream sequences, exon 1, the first intron, fused exons 2-5 and 30bp of 3' flanking sequences past the main polyadenylation signal [393]. In transgenic mice expression of this construct was extremely low and not tissue specific [393]. In contrast, a similar construct carrying an additional 5.6kb of 3' flanking sequences (now called the hCD2 minigene) was expressed in a tissue specific, copy number dependent manner, albeit at levels approximately 1/5 of that observed with constructs containing all of the introns (figure 1C) [393]. It was therefore proposed that an essential regulatory element is present within the 3' sequences downstream of the hCD2 polyadenylation signal. When these 3' sequences were used in combination with either the Thy-1 and β-Globin genes they were able to confer tissue-specific copy number-dependent expression on the transgene [393]. Interestingly, this pattern of expression was only noted in the thymocytes; whereas in other tissues (such as brain) transgene expression was subject to PEV [393]. These data indicate that the hCD2 LCR is contained exclusively within the 5.6kb 3' fragment and that the LCR functions in a tissue specific manner. Further deletional analysis has since revealed that the LCR lies within a smaller 2.1kb fragment immediately 3' of the polyadenylation signal (now called hCD2 2.1kb LCR. figure 1D)

[394]. When the LCR is linked to hCD2 transgenes position independent expression is always obtained, with the notable exception of X chromosome integrations where in females random X inactivation results in hCD2 transgene silencing in approximately 50% of T cells (Richard Festenstein unpublished data).

Detailed DNase I hypersensitivity assays have identified 3 clusters of HSS within the 2.1kb LCR (figure 1D) [341]. In transient transfection assays HSS1 functions as a classical enhancer, whereas, HSS2 and HSS3 have no enhancer activity at all [395]. However truncated hCD2 transgenes lacking HSS3 are sensitive to PEV, resulting in variegated hCD2 expression, indicating that although HSS3 is not an enhancer it is required for full function of the LCR in transgenic mice (figure 1E) [341]. Sorted $hCD2^+$ T cells from variegating lines display both HSS1 and HSS2 as well as a HSS corresponding to the hCD2 promoter, revealing that in these cells the transgene is in an open chromatin configuration [341]. In contrast, $hCD2^-$ T cells sorted from the same mice do not display either LCR HSSs or the promoter HSS, suggesting that that in these cells the transgene is in a closed chromatin configuration. Finally, FISH analysis has revealed that variegating transgenes are frequently integrated near centromeres or telomeres [341, 342]. Conversely, non-variegating truncated transgenes are always integrated in the chromosomal long arm [341, 342]. From these results it has been proposed that when hCD2 transgenes are integrated in areas rich in heterochromatin (such as near centromeres, telomeres or at certain sites within the long arm) then HSS3 is required to form an open chromatin domain permissive for gene expression in all cells. In the absence of HSS3 heterochromatin is able to spread from neighbouring regions into the transgene resulting in silencing in a proportion of cells. When HSS3 is present spreading of heterochromatin into the transgene is prevented and hCD2 is

expressed in all T cells. In contrast, if a truncated transgene is integrated at a site where the neighbouring chromatin is already in an open chromatin configuration (such as the chromosomal long arm which is mainly euchromatic) then HSS3 is not absolutely required to ensure expression of hCD2 in all T cells. Perhaps relevant to this notion the murine CD2 (mCD2) contains sequences that are homologous to the hCD2 HSS1 (transcriptional enhancer) but the homology does not extend further downstream to include HSS3 of the LCR, which may reflect the fact that mCD2 is located in the chromosomal long arm and not threatened by heterochromatin whereas hCD2 is pericentromeric.

1.12.5 Modifiers of hCD2 Variegation

Like PEV in *Drosophila*, variegation of hCD2 transgenes is sensitive to the concentration of certain chromatin regulatory proteins. For example, Festenstein et al. have shown that variegation of hCD2 transgenes carrying the truncated 1.3kb LCR is sensitive to the concentration of the murine HP1 homologue M31 [342]. Interestingly, over expression of M31 has two contrasting effects in these mice, which are dependant upon the chromosomal context; M31 over expression enhanced PEV in lines with centromeric or pericentromeric integrations, but decreased PEV of transgenes at non-centromeric integrations [342]. The mechanisms behind these two divergent phenotypes have not yet been characterised. In a complementary experiment, knocking out both M31 alleles resulted in decreased variegation of a pericentromeric hCD2 transgene (R. Aucott, submitted).

Analogous studies have revealed a similar dose dependant influence of proteins that are directly involved in the regulation of hCD2 expression. Fine mapping of the hCD2 LCR

HSS3 has revealed the presence of 4 footprints (FT1-4) that correspond to protein binding sites ([396] and T. Zhuma unpublished data). FT1 was subsequently shown to contain two HBP1 recognition sequences that are able to bind HBP1 in gel retardation assays [396]. HBP1 is an HMG box containing transcription factor that has been implicated in a number of biological events, such as cell cycle control and differentiation.

In transgenic mice deletion of the FT1 site disables the LCR resulting in PEV, indicating that the presence of FT1 is required for correct LCR function [396]. In a complementary experiment, variegating hCD2 transgenic mice containing only 1.5kb of LCR sequences, thus retaining the FT1 site, were generated. In these mice the extent of variegation was sensitive to the levels of HBP1 expression (Sekkali et al, in press). Increased levels of HBP1 in Lck-HBP1 transgenic mice resulted in a decrease in the level of hCD2 PEV (Sekkali et al, in press). In contrast, variegation in hCD2 transgenic mice containing only 1.3kb of LCR sequences, and therefore lacking FT1, was not influenced by over expression of HBP1 (Sekkali et al, in press). Together these results confirm that HBP1 plays a direct role in regulating hCD2 expression in transgenic mice.

1.12.6 Stability of Variegation

In most variegating mice the levels of hCD2 variegation are the same in the developing thymocytes and the mature peripheral T cells which suggests, but does not directly prove, that the decision to express or silence the transgene is stably maintained on subsequent cell division [397]. To address whether the decision to express hCD2 is stably maintained in the absence of HSS3, $hCD2^-$ and $hCD2^+$ populations were sorted from a variegating hCD2 transgenic line and stimulated to divide *in vitro* [397]. Flow

cytometry revealed that after 48 hours although 85% of the cultured cells had divided they still retained their original hCD2 phenotype [397]. Thus, *in vitro*, the decision to silence or express hCD2 appears to be stably maintained on cell division [397]. However, it is important to note that activating T cells can in some instances reduce hCD2 variegation (D. Kioussis unpublished observation).

Although these results confirm that HSS3 is required to ensure an open chromatin configuration when the transgene is integrated into heterochromatin, they do not address the role of HSS3 in maintaining an open chromatin structure after hCD2 expression has been initiated. Thus, the question to be answered is if HSS3 is ablated in a population of T cells that already express hCD2 does then transgene expression become sensitive to PEV, or is the former decision to express stably maintained even after the removal of HSS3? One potential approach to characterise the role of HSS3 in maintenance of LCR function would be to delete the site from hCD2 transgenes *in vivo* after the initiation of hCD2 expression, and then determine the effect of the deletion on transgene expression.

The main aims of this project were to:-

1. Use a number of techniques to identify novel regulatory elements within the hCD2 gene whose function could be addressed by deleting them using an *in vivo* Cre/loxP deletion strategy.

2. Generate a construct that could confer elevated levels of expression on hCD2 transgenes, without conferring protection against PEV in transgenic mice.

3. Address the influence of transgene copy number on hCD2 transgene variegation.

4. Generate transgenic Cre expressing lines to enable the future tissue-specific deletion of potential regulatory elements (such as HSS3) from the hCD2 gene.

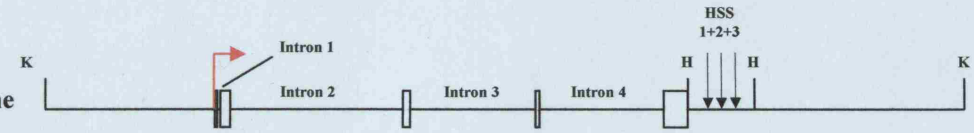
Figure 1. hCD2 expression constructs.

Open boxes represent exons and the positions of introns are indicated. The position of the 2.1kb LCR is indicated by two HinDIII sites which are separated by 2121bp of sequence containing the full LCR. The state of variegation and relative level of expression for each transgene is indicated to the left of each construct. The approximate locations of the three hypersensitive sites (HSS) within the hCD2 LCR are indicated. K= KpnI. H=HinDIII. X=XbaI.

A. hCD2 Maxigene

No Variegation

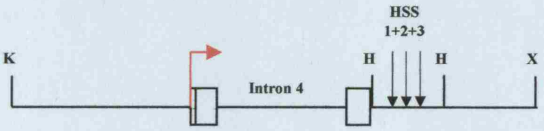
High Expression



B. hCD2-Int4

No Variegation

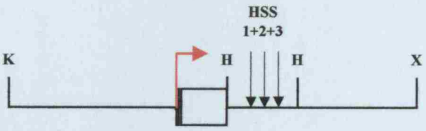
High Expression



C. hCD2 Minigene

No Variegation

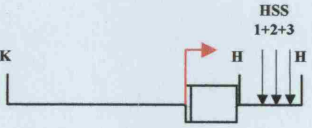
Low Expression



D. hCD2 2.1 kb LCR

No Variegation

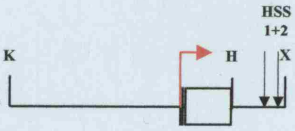
Low Expression



E. hCD2 1.3 kb LCR

Variegation

Low Expression



1kb

Chapter Two

Materials and Methods

2.1 Chemicals and Reagents

Chemicals were obtained from Sigma Aldrich Company and VWR International Ltd. Hybond-N+ nylon membranes, [α^{32}P] UTP and [α^{32}P] dCTP were obtained from Amersham Biosciences. Nitrocellulose membranes for slot blotting were obtained from Schleicher & Schuell. Kodak film was obtained from Anachem Ltd. Nick labelling kits were purchased from Invitrogen. Restriction endonucleases, T4 polynucleotide ligase, Klenow DNA polymerase, Proteinase K, DNase I and RNase A were obtained from Boehringer Mannheim and New England Biolabs. Custom-made oligonucleotides were obtained from Sigma Genosys. Tissue culture plastics were obtained from Marathon Laboratory Supplies and VWR International. Tissue culture media and reagents were obtained from Sigma Aldrich Company. FCS was obtained from GlobePharm.

2.2 Molecular Biology

2.2.1 Bacteriological Cultures

Liquid cultures were grown in L-Broth supplemented with 100μg/ml ampicillin. For agar plates, L-Broth was supplemented with 1.5% (w/v) bactoagar. Bacteria were frozen in 1 ml aliquots of L-Broth supplemented with 1/10 volume of 10x Hogness.

L-Broth:

1% (w/v) bactotryptone; 0.5% (w/v) yeast extract; 1% (w/v) NaCl in ddH_2O.

Hogness (10x):

36mM KH_2PO_4; 13mM K_2HPO_4; 20mM Sodium Citrate; 10mM $MgSO_4$; 40% (w/v) Glycerol.

2.2.2 Competent Bacteria and Cell Transformation

For competent bacteria, a single colony of *Escherichia coli* DH5α bacteria was picked from an agar plate and grown (250rpm in a New Branswicki Model G25 orbital shaker) overnight to stationary phase at 37°C in 2.5ml of L-Broth. 250ml of L-Broth with 20mM Magnesium sulphate was inoculated with 2.5 ml of the overnight culture and grown with shaking at 37°C until an optical density of 0.5 at 600nm was reached. Bacteria were then pelleted at 1285g, 4°C for 20 minutes and gently resuspended in 100ml of ice cold TFB1. Cells were left on ice for 5 minutes. Bacteria were then pelleted at 1285g, 4°C for 20 minutes and then re-suspended in 10 ml of ice cold TFB2. Cells were incubated for 45 minutes on ice and then frozen in 100µl aliquots in sterile Eppendorf tubes in a dry ice/methanol bath and stored at -70°C.

For transformation of bacteria, a ligation reaction containing ~100ng of DNA was mixed with 100µl of competent cells. Following incubation on ice for 30 minutes, the cells were heated to 42°C for 30 seconds and placed back on ice for another 2 minutes. 1 ml of L-Broth was added and cells were grown for 45 minutes at 37°C. 100µl of this culture was then plated out onto L-agar plates supplemented with the appropriate antibiotics.

TFB1:

30mM potassium acetate, 10mM $CaCl_2$, 50mM $MnCl_2$, 100mM RbCl 15% glycerol. pH adjusted to 5.8 with 1M acetic acid.

TFB2:

10mM MOPS (pH 6.5), 75mM $CaCl_2$, 10mM RbCl 15% glycerol. pH adjusted to 6.5 with 1M KOH.

2.2.3 Small-Scale Plasmid DNA Isolation

1.5 ml of a 5ml overnight culture of L-Broth supplemented with antibiotics was transferred to an Eppendorf tube and pelleted by centrifugation at 17900xg for 1 minute. The bacterial pellet was resuspended in 200µl STET and 20µl 10mg/ml lysozyme. The mixture was heated to 100°C for 1 minute. Debris was pelleted for 10 minutes at 17900xg in a bench top centrifuge. The pellet was removed and the plasmid DNA was precipitated from the supernatant with 600µl of 100% ethanol at 4°C and collected by centrifugation at 17900xg. The DNA pellet was resuspended in 100µl ddH_2O and stored at -20°C.

STET buffer:

0.1 M NaCl; 10 mM Tris-Cl, pH 8.0; 1 mM EDTA, pH 8.0; 5% Triton X-100

2.2.4 Large-Scale Plasmid DNA Isolation

Bacteria were harvested from a 1 litre over night culture by centrifugation at 3000xg for 30 minutes, at 4°C. The pellet was resuspended by vortexing in 40ml 1x glucomix and incubated for 5 minutes at RT. Bacteria were then lysed with 80ml of Alkaline Lysis Mix for 10 minutes and chromosomal DNA and protein was precipitated for 10 minutes on ice by the addition of 40ml 5M KOAc. Debris was removed by centrifugation for 20 minutes at 3000xg. The supernatant was filtered through two layers of cheesecloth and plasmid DNA precipitated at RT by the addition of 0.6 volumes of isopropanol and harvested by centrifugation for 20 minutes at 3000xg. The pellet was dissolved in 9ml

of ddH_2O. 10g of CsCl and 0.5ml of 10mg/ml ethidium bromide were added to the DNA solution. Samples were then transferred into Quickseal Ultracentrifuge Tubes (Beckman) and centrifuged at 275444xg for 15 hours in a Beckman VTi65.2 rotor. Banded plasmid DNA was recovered using a hypodermic needle and syringe and ethidium bromide was removed by repeated extraction using butanol saturated with ddH_2O. The DNA solution was then diluted with 3 volumes of ddH_2O and DNA was precipitated with 2.5 volumes of ethanol. After centrifugation for 30 minutes at 10000xg the DNA pellet was resuspended in 1ml ddH_2O. RNase A was added to the DNA to a final concentration of 20mg/ml and it was incubated at 37°C for 30 minutes. The DNA was then extracted once with equal volume of phenol:chloroform:isoamylalcohol 25:24:1 and precipitated by adding 3M NaOAc to 0.3M final concentration and 2.5 volumes of ethanol. The DNA pellet was washed with 70% ethanol, resuspended in 1ml ddH_2O and stored at –20°C.

Glucomix (10x):

250mM Tris-HCl, pH8.0; 500mM Glucose; 100mM EDTA.

Alkaline Lysis Mix:

200mM NaOH; 1% (w/v) SDS.

5M KOAc, pH4.8:

3M with respect to potassium; 5M to acetate; adjusted to pH4.8 with glacial acetic acid.

2.2.5 DNA Restriction Digests

Restriction digests were carried out according to the manufacturer's directions. DNA was incubated under the appropriate conditions for the restriction enzyme used. For digests with more than one enzyme, which had different buffer requirements, DNA was digested in consecutive single digests. The DNA was phenol:chloroform extracted and ethanol precipitated between these digests. For enzymes that required the same restriction buffer, both enzymes were used simultaneously in the same reaction.

2.2.5 Agarose Gel Electrophoresis

Agarose gels were prepared in 1x TAE and run horizontally submerged in 1x TAE buffer, supplemented with 5μg/ml of ethidium bromide. DNA samples were run with 1:5 volume orange G loading dye and loaded in parallel with a 1kb ladder (Invitrogen). DNA was visualized and photographed under UV light.

50x TAE:

2M Tris-Acetate; 100mM EDTA; pH8.0.

Orange G loading dye:

20% ficoll; 110mM Tris-HCl, pH7.5; 0.25% (w/v) orange G dye.

2.2.7 Extraction of DNA from Agarose Gels

Following DNA digest and electrophoresis, the gel was viewed under UV light and the appropriate band cut out of the gel. The gel was snap frozen in liquid nitrogen and gradually defrosted. The resultant liquid, containing the DNA, was phenol extracted (without chloroform in the initial stage) and precipitated with ethanol.

2.2.8 Sequencing

50ng of purified plasmid DNA was sequenced with 3.2 pmoles of the appropriate primer and 8µl of Perkin Elmer ABI Prism Big Dye Terminator cycling reaction in a final volume of 20µl according to manufacturer's instructions. Cycle sequencing was carried out using a Hybaid MBS PCR machine.

Reaction conditions:

25x [96°C 30 sec. 50°C 15 sec. 60°C 4 min]

Extension products were precipitated and simultaneously purified from unincorporated Dye-terminators by ethanol precipitation at room temperature for 15 minutes followed by centrifugation at 17900xg for 20 minutes. Pelleted extension products were washed with 70% ethanol and spun for 5 minutes. The pellets were air dried and resuspended in 5µl of loading buffer.

Sequence data was acquired on an Applied Biosystems (ABI) Model 377 Automated DNA sequencer with 36 cm well to read plates and 5% poly-acrylamide gels using standard running conditions with a 9 hour collection time. Sequence data was analyzed and edited with ABI Sequence Analysis. Edited DNA sequence was further analysed using Redasoft visual cloning and Blast software form NCBI.

5% Polyacrylamide gel:

5 ml 10x TBE; 5ml Long ranger acrylamide solution, FMC Bioproducts; 18 g of urea; ddH_2O to 50ml; 250µl of 10% (w/v) APS; 25µl TEMED

Loading buffer:

5 parts deionised formamide; 1 part 25 mM EDTA (pH 8.0) containing 50 mg/ml Blue dextran.

2.2.9 Oligo Annealing

Complementary oligos were resuspended at equimolar quantities in annealing buffer. The oligos were heated in a boiling water bath for 10 minutes and then allowed to slowly cool to room temperature.

Annealing Buffer:

10mM Tris, pH 8.0, 50mM NaCl, 1mM EDTA

2.3 Analysis of Genomic DNA

2.3.1 Genomic DNA Preparation

Tissue samples were digested at 56°C overnight in lysis buffer and 200μg/ml proteinase K in shaking water bath. Digested samples were centrifuged (14,000xg for 15 minutes) and the supernatant transferred to a fresh tube. One volume of phenol/chloroform was added to the sample and the sample underwent shaking for 15 minutes. The sample was centrifuged (14,000xg for 15 minutes) and the supernatant transferred to a fresh tube. Extraction with phenol was repeated until the protein/DNA interface was clear. Samples were then mixed with one volume chloroform and centrifuged as before. The resultant supernatant was transferred to a fresh tube and DNA precipitated by addition of one volume isopropanol. Precipitated DNA was spooled out with a sealed Pasteur pipette,

washed in 70% ethanol and air-dried. DNA was resuspended in 50-100µl H_2O. The concentration of DNA was determined by measurement of absorbance of the sample at 260nm and 280nm on a spectrophotometer and concentration calculated according to the following formula.

$$[\text{DNA}]\ \mu g./\mu l = \frac{\text{Abs 260 x 50 x dilution}}{1000}$$

The ratio of Abs260: Abs280 was calculated to determine the quality of DNA.

Lysis buffer:

100mM Tris HCl, pH 8.5; 5mM EDTA; 0.2% SDS; 200mM NaCl.

Phenol/chloroform:

phenol:chloroform:isoamyl alcohol (25:24:1 v/v) equilibrated with 10mM Tris-HCl, pH7.5.

2.3.2 Slot Blot Analysis

10µg of genomic tail DNA was denatured in 400µl of 0.4M NaOH for 10 minutes before neutralisation by the addition of 400µl of 2M Ammonium acetate. A Schleicher and Schuell slot blot apparatus was assembled according to the manufacturer's instructions using a Schleicher and Schuell nitrocellulose membrane pre-wet in 1M Ammonium acetate. The apparatus was connected to a water vacuum before loading samples in duplicate using 400µl of each sample per slot. After loading the filter was removed and the DNA fixed by baking at 80°C for 1 hour.

2.3.3 Southern Blot Analysis

10μg (unless otherwise stated) of genomic DNA was digested in a final volume of 40μl overnight in the appropriate conditions for the restriction enzyme using 40 units of enzyme. To check completion of digestion, 3μl of this reaction was mixed with 0.5μg of lambda DNA in a second tube and incubated in parallel. Digests were subject to electrophoresis on a 0.7-1% agarose gel run at (2V/cm) overnight. After denaturation with blot solution I for 40 minutes and subsequent neutralisation in blot solution II for 40 minutes, separated DNA fragments were transferred onto a Hybond-N+ nylon membrane (Amersham Biosciences), essentially as describe in ref [398]. DNA was then fixed onto the membrane by baking at 80°C for 2 hours.

Blot solution I:

0.5M NaOH; 1.5M NaCl.

Blot solution II:

0.5M Tris-HCl, pH7.5; 3M NaCl.

2.3.4 DNA Probe Labelling

DNA probes were labelled by random priming using Ready-To-Go DNA labelling Beads (Amersham Biosciences). 50ng of DNA was resuspended in 47μl of H_2O, denatured at 100°C for 5 minutes and then immediately cooled on ice. Denatured DNA was added with 3μl of [α32P]dCTP (1.11MBq) to a Ready-To-Go DNA labelling Bead and incubated at 37°C for 15 minutes. Unincorporated nucleotides were removed using a Pharmacia G-50 nick column according to the manufacturer's instructions. The

collected probe was denatured by boiling at 100°C for 5 minutes with 400µl of denatured salmon sperm (10mg/ml) before use in hybridisation reactions.

2.3.5 Filter Hybridisation

Filter membranes were pre-hybridised at 65°C for at least 2 hours in 50ml of hybridisation buffer in a Hybaid bottle placed in a rotating Hybaid oven. 50ng of labelled denatured probe was added to the hybridisation buffer and left to hybridise overnight at 65°C. Following hybridisation, filters were washed in a final stringency of 0.3x SSC; 0.1% (w/v) SDS at 65°C changing wash solutions every 20 minutes. Filters were then prepared for autoradiography.

SSC (20x):

3M NaCl; 0.3M NaCitrate.

Hybridisation buffer:

10x Denhardts; 3x SSC; 0.1% (w/v) SDS; 10% (w/v) Dextran Sulphate; 100µg/ml denatured salmon sperm DNA.

100x Denhardts:

2% (w/v) BSA; 2% (w/v) Ficoll 400; 2% (w/v) Polyvinyl pyrollidone.

2.3.6 Autoradiography

Washed nylon membranes from hybridizations were wrapped in Saran Wrap and exposed to Kodak Biomax photographic film at -70°C using intensifying screens.

2.3.7 DNase I Assay

Preparation of nuclei from mouse tissue was adapted from Forrester et al. [399]. The thymus was dissected from culled mice and stored in PBS on ice. The thymus was carefully teased on ice, passed through a 70μm cell strainer and centrifuged (275xg at 4°C for 5 minutes). Cells were resuspended in PBS, counted (Schärfe System) and $1x10^7$ cells/sample aliquoted and centrifuged (275xg at 4°C for 5 minutes). The cell pellet was gently pipetted to resuspend, added to 5ml reticulocyte standard buffer (RSB) and 3ml 0.5% NP40 and incubated on ice for 5 minutes. 5 volumes of RSB were added to stop cell lysis and the cells centrifuged (275xg at 4°C for 7 minutes). The nuclei pellet was resuspended in 1ml RSB.

$1x10^7$ nuclei were added to 1μl 1M $CaCl_2$ in ice-cold DNase I free eppendorf tubes and kept on ice. The first sample was incubated in the absence of DNase I, while remaining samples were incubated with increasing amounts of DNase I (20ng, 40ng, 80ng, 100ng, 200ng, 300ng, 600ng, 800ng and 1μg from 4mg/ml stock) at 37°C for 4 minutes. Following incubation 100μl DNase I stop mix was added to inhibit DNase I activity. Samples were then digested with 100μg proteinase K at 56°C overnight with shaking. Samples underwent DNA extraction, Southern blot analysis, hybridisation and autoradiography as previously described.

RSB (10x):

100mM Tris-HCl, pH7.5; 100mM NaCl; 30mM $MgCl_2$.

0.5% NP40 solution:

0.5% NP40 (w/v); 1X RSB.

DNase I stop mix:

0.6M NaCl; 20mM Tris, pH 8.0; 10mM EDTA; 1% (w/v) SDS.

2.3.8 McrBC Titration Assay

The MCrBC assay was adapted from Santoso et al. [400]. 50μg of genomic DNA was digested over night with the restriction enzyme HindIII, which generates a hCD2 LCR fragment of 2.1kb. After digestion, the DNA was ethanol precipitated and resuspended in 55μl of 10mM Tris (pH 8.0). The DNA samples were divided into five 10μl aliquots and digested as recommended by the manufacturers with increasing amounts (0-40 enzyme units) of the DNA methylation-dependent McrBC restriction enzyme (New England Biolabs) for 20min at 37°C in 20μl final volume. The reaction was stopped by the addition of 2μl of stop solution. Samples were resolved on 1% agarose gels containing 1x TAE and ethidium bromide (5μg/ml), run at 2 volts/cm overnight, and analyzed by Southern blot as previously described.

Stop solution (10xTES):

100mM Tris-Cl (pH7.5); 10mM EDTA (pH7.5); 1% (w/v) SDS.

2.4 Analysis of RNA

2.4.1 RNA Isolation from Mouse Tissues

Mouse tissue was homogenised in 3ml of 3M LiCl, 6M urea with an ultra turrex for 1 minute. Samples were sonicated for 50 seconds to shear the genomic DNA and left over night at 4°C to precipitate the RNA. RNA was pelleted by centrifugation at 10,000xg at

4°C for 40 minutes. The RNA was washed with 0.5ml of 3M LiCl, 6M urea and re-pelleted by centrifugation. RNA was resuspended in 0.5ml 10mM Tris pH7.5 1mM EDTA. 0.5% SDS and extracted twice with an equal volume of Phenol/chloroform. RNA was precipitated with 1 volume of isopropanol and washed once with 500µl of 75% EtOH. The RNA pellet was dissolved in 100µl of ddH_2O and stored at -70°C.

2.4.2 RNA Isolation from Cell Lines

RNA was isolated from 10-30x10^6 cells using TRIZOL (GIBCOBRL. LIFE TECHNOLOGIES) as recommended by the manufacturers. RNA was resuspended in ddH_2O and stored at -70°C.

2.4.3 RNase Protection Analysis

RNase protection was performed essentially as recommended by the manufacturers (RiboQuant Ribonuclease protection assay kit). The hCD2 LCR pBluescript II SK + and hCD2 LCR pBluescript II KS + vectors were linearised with BamHI and HinDIII respectively, cleaned by phenol/chloroform extraction and ethanol precipitated. Linearised vectors were resuspended to give a final concentration of 50ng/µl. 1µl of each linearised vector was used in the probe synthesis step in place of the RPA template set provided by the manufacturers. In addition, only 3µl of [α32P]UTP was used rather than the 10µl recommended by the manufacturers. All subsequent stages were performed as recommended by the manufacturers. Protected probes were resolved by electrophoresis on a 6% polyacrylamide gel at 50 Watts for 2-3 hours. Resulting gels were dried under vacuum and exposed to x-ray film at -70°C.

2.4.4 Nested RT-PCR

For generation of cDNA the Gene Amp PCR core kit was used as recommended by the manufacturers (Roche).

Primers used for cDNA generation:

RT-AS GACTAGACCCGTGTTGC

RT-S GGAGTTCAGAAGAAAGGC

Reaction conditions:

1X [42°C 14min] 1X [99°C 5min] 1X [5°C 5min]

For PCR reactions 1µl of each cDNA was used per PCR reaction in a final volume of 25µl. A 2x PCR master mix (2x ReddyMix PCR Master Mix. ABgene) was used to minimise pipetting errors. 50pmol of each PCR primer was used per reaction (final concentration of each primer is 2µM).

PCR primers:

F1 GCTGGTGCCTTGTCTGTG

R1 CAATTATTCTGCCTCAACC

PCR reaction conditions:

1x [94°C 3min] 30x [94°C 40sec. 52°C 40sec. 72°C 30sec] 1x [72°C 5min] 1x [4°C hold]

Samples were resolved on 2% agarose gels containing 1x TAE and ethidium bromide (5µg/ml)

2.4.5 Non-Coding RT-PCR

Reverse Transcriptase Reaction:

Reaction Buffer (AmpliTaq GOLD buffer II)	1.5µl
$MgCl_2$ (25mM)	2µl
dNTPs (10mM Each)	1.5µl
RT Primer (12.5µM)	0.16µl
RNasin	1µl
MMLV	0.7µl
RNA	1µl
H_20	7.14µl

All reagents are from Perkin Elmer

RT Primer:

TCCCGATTAAGCCATACCACGAAGTGACACGATCTCACCT

Reaction conditions:

1x [37°C 60min] 1x [95°C 5min] 1x [10°C Hold]

The generated cDNA was purified using a Quiaquick PCR purification column as recommended by the manufactures, before being used for PCR.

PCR Reaction:

Reaction Buffer (AmpliTaq GOLD buffer II)	2µl
$MgCl_2$ (25mM)	1.6µl
dNTPs (10mM Each)	2µl
Forward Primer (25µM)	0.2µl
Reverse Primer (25µM)	0.2µl
Taq (AmpliTaq GOLD)	0.1µl
Purified cDNA	2µl
H_20	11.9µl

All reagents are from Perkin Elmer

Forward Primer:

CCAGCTAGGTAACAATTGCC

Reverse Primer:

TCCCGATTAAGCCATACCAC

Reaction Conditions

1x [95°C 10min] 30x [94°C 30sec. 60°C 45sec. 72°C 1min] 1x [72°C 10min] 1x [4°C hold]

Samples were resolved on 2% agarose gels containing 1x TAE and ethidium bromide (5µg/ml).

2.5 Genotyping of Transgenic Mice by PCR

2.5.1 PCR Genotyping of Cre Recombinase Expressing Mice

PCR reaction:

Taq (Abgene thermoprime plus)	0.3µl
Buffer (Abgene 10X reaction buffer IV)	2.0µl
$MgCl_2$ (25mM)	1.6µl
dNTPs (10mM each)	0.4µl
Primer 1 (50µM)	0.1µl
Primer 2 (50µM)	0.1µl
DNA	1.0µl
H_20	14.5µl

Reaction Conditions

1x [95°C 3min] 30x [94°C 40sec. 62°C 40sec. 72°C 30sec] 1x [72°C 5min] 1x [4°C hold]

2.5.2 PC3-Cre and CD19-Cre genotyping primers.

Forward Primer:

TTCCCGCAGAACCTGAAGATGTTCG

Reverse Primer:

GCCAGATTACGTATATCCTGGCAGC

2.5.3 Vav-iCre and hCD2-iCre Genotyping Primers.

Forward Primer:

AGATGCCAGGACATCAGGAACCTG

Reverse Primer:

ATCAGCCACACCAGACACAGAGATC

Samples were resolved on 2% agarose gels containing 1x TAE and ethidium bromide (5μg/ml).

2.5.4 PCR Genotyping of ROSA26 knockin Mice (R26R-EYFP and DIP-R1)

PCR reaction:

Reaction buffer (Roche diagnostics 10x)	2.0μl
$MgCl_2$ (25mM)	1.6μl
dNTPs (10mM each)	0.4μl
Forward Primer (50μM)	0.1μl
Reverse Primer (50μM)	0.1μl
Glycerol (6.75%)	7.3μl
Taq polymerase	0.3μl
DNA sample	1.0μl
H_2O	7.4μl

Reaction Conditions

1x [94°C 2min. 55°C 1min] 29x [72°C 45sec. 94°C 30sec. 55°C 45sec] 1x [72°C 3min. 20°C 30sec] 1x [4°C hold].

Forward Primer:

AAG TCG CTC TGA GTT GTT AT

Reverse Primer:

GCG AAG AGT TTG TCC TCA ACC

Samples were resolved on 2% agarose gels containing 1x TAE and ethidium bromide (5μg/ml).

2.6 Gene Targeting

2.6.1 Expansion and Trypsinisation of ES Cells

R1 embryonic stem cells from 129/Sv mice were maintained in the undifferentiated state by growth on a feeder layer of γ-irradiated primary embryonic fibroblasts using culture medium supplemented with leukaemia inhibitory factor (LIF). The ES cells were grown to near confluency on 6cm culture dishes in 5ml of ES medium. The culture dishes were pre-coated with 0.1% Gelatine for 30 minutes and contained a monolayer of $2x10^6$ embryonic fibroblasts. The culture medium was changed every day and when the cells reached confluency, they were washed with 5 ml of PBS and 0.5 ml of trypsin was added to them. The cells were incubated at 37°C for 3 minutes and then pipetted up and down vigorously with a Pasteur pipette to ensure that a single cell suspension was obtained. Cells were spun through medium, resuspended in 5 ml of ES medium and 1/10 of them were transferred to a fresh 6 cm culture dish with a feeder layer.

ES Medium:

Dulbecco's Modified Eagle's Medium (DMEM) supplemented with; 1mM Sodium Pyruvate; 15% Foetal calf serum (FCS); 2mM Glutamine; 5U/ml Penicillin; 50μg/ml Streptomycin; 0.1mM Non-essential amino acid (MEM); 100μM β-mercaptoethanol; 1000U/ml Leukaemia inhibitory factor (LIF).

Phosphate buffered saline (PBS):

1.4M NaCl; 27mM KCl; 81mM Na_2HPO_4; 15mM KH_2PO_4; adjusted to pH 7.5 with HCl

Trypsin:

0.25% Trypsin powder; 1.4mM EDTA; 120mM NaCl; 790μM $Na_2HPO_4.12H_2O$; 1.4mM KH_2PO_4; 5mM KCl; 5.5mM D-glucose; 16.5 Tris; 14μM Phenol red; adjusted to pH7.6 and filtered.

2.6.2 Preparation of Embryonic Fibroblasts

To make G418 resistant embryonic fibroblasts (EFneo), a $\beta 2m^{-/-}$ or an $IL3^{-/-}$ (both of which contain the Neo gene) male mouse was mated to a wild-type female. The pregnant female was sacrificed at day 13 or 14 of gestation and the embryos were dissected out free of extra-embryonic membranes in sterile conditions. Each embryo was washed in 70% ethanol and subsequently in PBS and soft organs and viscera (liver, gut, and brain) were dissected out. The remaining carcass was cut into small pieces and placed into 2ml of trypsin in 15ml tube (FALCON). The tube was left at 4°C for overnight and then incubated at 37°C for 10 minutes, with occasional pipetting, before the addition of an equal volume of EF medium. The big pieces of embryonic tissue were left to settle for 2 minutes, the supernatant was removed and plated out in one $175cm^2$ tissue culture flask in 50ml of EF medium. The flask was incubated at 37°C and 5% CO_2. The medium was changed the next day and the fibroblasts were left to grow for 2 further days until they reached extreme confluency. The fibroblasts from one flask were split 1/5 and passed into 5 new $175cm^2$ flasks. When they reached confluency after 3 days, the fibroblasts from each flask were split 1/5 and were passed into five $175cm^2$ flasks again. They were left to grow for further 3 days and then the fibroblasts from all

25 flasks were collected, γ-irradiated (3500 rads) and frozen in EF medium containing 12% DMSO at $4x10^6$ cells/vial or $20x10^6$ cells/vial. The total yield of fibroblasts is 400-$600x10^6$ irradiated EFs/embryo. The irradiated EFs were used to form a monolayer on the culture dish in which the ES cells were grown. One vial of $4x10^6$ cells is enough for 1x10cm plate or 2x6cm plates.

EF Medium:

Dulbecco's Modified Eagle's Medium (DMEM) with; 1mM Sodium Pyruvate; 15% Foetal calf serum (FCS); 2mM Glutamine; 5U/ml Penicillin; 50μg/ml Streptomycin; 0.1mM Non-essential amino acid (MEM); 100μM β-mercaptoethanol.

2.6.3 Transfection and Selection of ES Cells

The ES cells were grown to 50-100% confluency. The DNA used for the transfection was a CsCl banded plasmid, linearised, ethanol precipitated and redisolved in sterile TE at concentration 2mg/ml. 10cm and 6cm culture dishes with $4x10^6$ and $2x10^6$ embryonic fibroblasts respectively were prepared in advance to receive the electroporated cells. The ES cells were trypsinised, washed with medium and resuspended in Hepes buffered saline (HBS) containing 0.0007% 2-mercaptoethanol (2-ME) at a concentration of 10^7cells/ml. The linearised targeting construct was added to the stock of cells at a final concentration of 50μg/ml and 0.8ml of this mix was used for each electroporation. The electroporations were done with a BIORAD Gene Pulser at 400Volts and 25μF using 0.4cm electrode gap micro-cuvettes (BIORAD). The cells were transferred then to a 15 ml Falcon tube containing 3.4 ml medium and plated out onto 4x10cm plates (0.9ml/plate). Selection was initiated 24 hours later at a concentration of 300μg/ml G418.

The selective medium was changed every 3 days and resistant colonies were picked after 10 days into 96-well plates, in 20µl PBS using P-200 Gilson pipette and tip. 6 days later the clones were transferred to 24-well plates and expanded for another 5 days. At confluency, half of each well was frozen in ES medium containing 12% DMSO and the other half was used to make DNA.

Hepes buffered saline (HBS):

137mM NaCl; 5mM KCl; 0.7mM Na_2HP0_4; 6mM dextrose; 21mM Hepes, pH7.1.

G418 stock:

50mg/ml in PBS stored up to a month at 4°C.

Gancyclovir stock:

1mg/ml in PBS stored frozen.

2.6.4 Karyotyping of ES Cells

ES cells were grown to 50-75% confluency, the medium was changed and 2 hours later Demecolcine (Sigma) was added at a final concentration of 20ng/ml to arrest the cells in metaphase. Six hours later the cells were trypsinised for 3 minutes at 37°C, washed with ES medium and washed once in PBS. They were resuspended in 5ml KCl 0.56% (Hypotonic solution) and let stand at room temperature for 6 minutes followed by centrifugation at 100xg for 5 minutes. The pellet was resuspended carefully in 5ml ice cold fixative by adding it dropwise and flicking thoroughly, left at room temperature for 5 minutes and spun down for a further 5 minutes. The fixing procedure was repeated

another 3 times and finally the cells were resuspended in 1ml fixative and dropped from 30cm height on a microscope slide that had been washed in 5% Acetic Acid in ethanol. The chromosome spreads were stained with 10% Giemsa's (DBH) for 20 minutes and chromosomes were counted using a high powered microscope.

Fixative:

Methanol: Acetic Acid 3:1

2.7 General Tissue Culture

2.7.1 HEK293 Cells

HEK293 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS) and 2mM L-glutamine. Cells were incubated at 37°C in a 5% CO_2 atmosphere and passaged when cells reached confluence of 70-80%.

2.7.2 Jurkat Cells

Jurkat cells were maintained in RPMI 1640 supplemented with 10% FBS and 2mM L-glutamine. Cells were incubated at 37°C in a 5% CO_2 atmosphere and split 1:3 approximately every 3 days.

2.8 Generation of Transgenic and Knockin Mice

2.8.1 Transgenesis

Before microinjection of DNA constructs, the prokaryotic sequences were removed by restriction enzyme digestion followed by a purification of the construct using agarose gel electrophoresis as described before. Isolated fragments were further purified using Elutip columns. Pronuclei of fertilised oocytes from (CBA/Ca X C57F7BL/10) Fl mice were injected with the purified DNA at a concentration of 1-2μg/ml TE buffer. Eggs surviving microinjection were transferred into the oviducts of recipient pseudopregnant females. Transgenic founders were detected by Southern blot analysis of genomic tail DNA and/or flow cytometry of blood. Pronuclei injections and transfers were carried out by Dimitris Kioussis and Mauro Tolaini.

2.8.2 Injection of ES Cells into Blastocysts and Generation of Knockin Mice

The cells were grown to 70-80% confluency, then trypsinised for 3 minutes at 37°C, washed once with ES medium and resuspended in 5ml injection medium. 3.5-day blastocysts were collected from C57Bl/6 females and kept in blastocyst medium. 10-15 ES cells were injected into each blastocyst and 10-12 blastocysts were transferred into the uterus of each foster mother. Chimeric males were bred to C57Bl/6 females and germline transmission was determined by agouti coat colour. Blastocyst injections and uterine transfers were carried out by Dimitris Kioussis and Mauro Tolaini.

Injection medium:

Dulbecco's modification of Eagle's medium with 20mM Hepes, without glutamine and bicarbonate (ICN Biomedicals); 10% Foetal calf serum; 2mM Glutamine; 5U/ml Penicillin; 50μg/ml Streptomycin.

Blastocyst medium:

Dulbecco's modification of Eagle's medium; 10% Foetal calf serum; 2mM Glutamine; 5U/ml Penicillin; 50μg/ml Streptomycin.

2.9 Mouse Work

2.9.1 Fluorescence Microscopy

Photographs were taken using a Nikon coolpix 5000 digital camera mounted on a Leica MZ FLIII fluorescence stereomicroscope fitted with a GFP fluorescence filter.

2.9.2 Isolation of Intestinal Epithelial Cells and Intraepithelial Lymphocytes

The small intestine was dissected out and the fecal matter expelled by flushing with 40 ml of cold CMF. Peyer's patches and adipose tissue were removed from the outer-surface of the intestine before dissecting into 0.5cm pieces.

Intestinal pieces were washed 3 times in 40ml cold CMF to remove any remaining fecal matter and mucus. The clean intestinal pieces where placed into 20ml of CMF/FCS/DTE and incubated at 37°C for 20 minutes with occasional vortexing. Intestinal pieces were removed by passing the supernatant through a 70μm cell strainer. Cells were pelleted by centrifugation (400xg for 5 minutes) and resuspended in PBS-azide for use in flow cytometry

CMF solution:

1x Hank's balanced salt solution (HBSS) Ca^{2+} and Mg^{2+} free; 1x HEPES-bicarbonate buffer; 20% FCS.

CMF/FCS/DTE:

CMF supplemented with 10% FCS and 1mM dithioerythritol

10x HBSS:

1.37M NaCl; 53mM KCl; 55mM glucose; 4.4mM KH_2PO_4; 3.3mM Na_2HPO_4; 400μM phenol red.

10x HEPES-bicarbonate buffer

100mM HEPES; 250mM sodium bicarbonate; adjusted to pH 7.2 with HCL.

2.9.3 Isolation and Enrichment of Double Negative Thymocytes.

To enrich for double negative thymocytes, $CD4^+$ and $CD8^+$ thymocytes were depleted by staining single cell suspensions from thymus with anti-CD4 (RL172.4) and anti-CD8 (AD4(15)) followed by treatment with guinea pig complement (Life Technologies) and rabbit complement (Sigma). Resulting viable cells were recovered on a ficoll gradient and used for flow cytometry.

2.9.4 Flow Cytometry

Single cell suspensions were prepared from thymus, lymph node, femoral bone marrow and small intestine. For flow cytometry $1X10^6$ cells were stained in 100μl ice-cold PBS supplemented with 0.5% BSA and 0.02% sodium azide. Cells were incubated for 30 minutes on ice in 96 well plates before washing twice with 100μl ice-cold PBS azide. Cells were again resuspended in 100μl PBS azide before data acquisition using a FACSCalibur and CELLQUEST software. All antibodies were used at appropriate

saturating concentrations. The antibodies used were as follows: anti-Ter119-Biotin (Pharmingen), anti-Mac1-APC (Pharmingen), anti-Gr1-Biotin (Pharmingen), anti-CD3-Biotin (Caltag), anti-CD19-APC (1D3), anti-IgD-Biotin (1.19), anti-CD4-PECY7 (Caltech), anti-CD8-APC (Caltech), anti-Thy1-PE (Pharmingen), anti-CD44-APC (Caltag), anti-CD25-Biotin (Pharmingen), Streptavidin-PECY7 (Caltag), anti-hCD2-FITC (Pharmingen), anti-CD4-PE (Pharmingen), anti-CD8-Tri colour (Caltag), anti-CD45-APC (Pharmingen), anti-TCRβ-FITC (eRioscience), anti-B220-Biotin (SouthernBiotech), streptavidin-PerCP-(Pharmingen).

2.9.5 Purification of T cells and B Cells from Spleen

Single cell suspensions were generated from spleen samples. Splenocytes were stained for 10 minutes at 4°C with saturating amounts of either biotinylated anti-CD3 or biotinylated anti-B220. After washing, cells were incubated for 10 minutes at 4°C with the appropriate number of streptavidin MicroBeads (Miltenyi Biotech). After washing the labelled cells were purified using LS separations columns (Miltenyi Biotech) according to the manufacturer's instructions. Genomic DNA was purified from positively selected cells essentially as described in 2.3.1.

Section One

Recombinase Based Technology in the Analysis of the hCD2 LCR

Experimental Rational

It is known that the HSS3 of the hCD2 LCR is required for the protection against PEV and the formation of an open chromatin configuration in transgenic mice [341]. However, it is not known whether HSS3 is required only for initiating the formation of an open chromatin configuration or whether it is required for the maintenance of the open chromatin configuration once formed. This led to the question of what would happen to hCD2 expression if HSS3 was deleted after hCD2 expression had been initiated? Would the hCD2 transgene become sensitive to PEV, resulting in the loss of hCD2 expression in a proportion of T cells? For this purpose, we set out to exploit the Cre/loxP recombination system, which has been proven to be an efficient tool for selective deletion of genomic sequences both *in vivo* and *in vitro* [401-403]. Cre, a P1 bacteriophage recombinase, is able to mediate intra- and inter-molecular recombination between pairs of direction-specific 34bp sequences termed loxP sites [404, 405]. Intra-molecular recombination between two loxP sites positioned in the same orientation results in excision of the intervening sequences, leaving a single loxP site. In contrast, recombination between loxP sites positioned in opposing orientations results in inversion of the intervening sequences and retention of both loxP sites. Transgenic models have shown that when Cre expressing transgenic mice are bred to mice containing genomic sequences flanked by loxP sites (conventionally termed floxed) Cre is able to excise the floxed target sequences from the genome [401, 406-416]. Cre can either be introduced into the genome by regular transgenesis or can be knocked into a defined locus to give the desired expression characteristics [401, 406-416].

The experimental strategy was to generate transgenic mice containing a single copy of a hCD2 transgene in which loxP sites flank HSS3 (or other regulatory elements identified within the hCD2 gene). Using Cre recombinase expressed conditionally under the control of different promoters HSS3 would be specifically deleted at different developmental time points (i.e. before or after initiation of hCD2 expression) and the effect on hCD2 expression analysed. One potential caveat of such an approach is that classical transgenesis frequently results in the integration of multiple transgene copies in a concatamer. Cre mediated recombination between loxP sites on adjacent copies of the hCD2 transgene would result in the deletion of entire copies of the transgene. Deletion of transgene copies may influence both the level of hCD2 expression and the pattern of variegation, thus, complicating interpretation of any results obtained from deletion of HSS3. Transgenic mice containing multiple transgene copies could therefore not be used in such an experimental approach. Thus, for this experimental strategy to be successful a number of tools had to be generated. First, a hCD2 transgene that could be expressed at detectable levels from a single copy was required. Second, transgenic lines expressing Cre at various developmental times points were needed. Finally, a mechanism for targeting a single copy of the hCD2 gene in a predetermined site of the genome is currently being developed.

Chapter Three

The Role of Intron Four and the Influence of Transgene Copy Number on Variegation.

3.1. Intron 4 and Position Effect Variegation

As mentioned above only single copy transgenic lines could be utilised in a Cre/loxP based approach to delete HSS3 in an inducible manner. One potential problem of this strategy is that the level of expression from a single copy of the hCD2 minigene (figure 1C) falls below the sensitivity of flow cytometry. In order to overcome this obstacle it was necessary to use a construct that could express hCD2 at higher levels. Previously it has been noted that inclusion of intron 4 in hCD2 transgenes (denoted hCD2-Int4 see figure 1B) results in significantly increased levels of expression (Richard Festenstein Ph.D. thesis). The exact mechanism by which these increased levels are obtained is undefined; however, inclusion of intronic sequences has long been known to increase transcription, mRNA stability, mRNA processing and nuclear export [417]. In addition, intron 4 contains a DNase I hypersensitive site (HSS), indicating that the intronic sequences may contain enhancer like elements (Richard Festenstein Ph.D. thesis). Transient transfection CAT assays indicate that this region has no enhancer function (D. Kioussis, unpublished observation). Importantly, however, the function of intron 4 has never been studied in the context of variegating hCD2 transgenic mice (i.e. in transgenes that lack HSS3). It is therefore not known whether inclusion of intron 4 (and the HSS that it contains) can compensate for the loss HSS3 from the LCR, thus protecting against PEV in transgenic mice.

To determine whether or not intron 4 can compensate for the absence of HSS3 and thus modify the variegation of a hCD2 transgene with a disabled LCR, the hCD2-int4 construct was modified by replacing the full LCR with a truncated 1.3kb LCR, thereby omitting HSS3. Finally, a single loxP site was included at the terminal 3' of the

construct to allow transgene copy number reduction following Cre mediated recombination.

3.2 Cloning

To generate the hCD2-Int4-1.3loxP vector pBluescript II SK + was cut with SpeI, blunted (using the DNA Polymerase I Large [Klenow] Fragment) and then cut with BamHI. The VA vector was cut with BamHI-XmnI and the 1.3kb LCR fragment isolated and ligated into pBluescript II SK+ to generate 1.3LCR-pBS. The 1.3kb LCR was lifted from the vector by digestion with BamHI-XbaI and ligated into the BamHI-XbaI digested hCD2-Int4 vector to generate hCD2-Int4-1.3. Finally, the hCD2-Int4-1.3 vector was digested with XbaI-NotI and a loxP site generated from annealed oligos ligated in (see below) to give the final construct hCD2-Int4-1.3-loxP (later shortened to 1.3loxP). See figure 2 for a schematic description of the cloning and figure 3 for a diagram of the final construct drawn to scale.

loxP sense oligo:

5'CTAGAATAACTTCGTATAATGTATGCTATACGAAGTTATGC3'

loxP anti-sense oligo:

5'GGCCGCATAACTTCGTATAGCATACATTATACGAAGTTATT3'

The final vector was linearised by digesting with KpnI-NotI and the hCD2 gene isolated from the plasmid sequences. The linearised construct was injected into the pronuclei of fertilised mouse oocytes before implantation into pseudopregnant foster mothers. The presence of the transgene in founder mice was confirmed by Southern blot analysis and

Figure 2. Generation of the hCD2-Int4-1.3-loxP construct.

The hCD2 minigene was cut with BamHI/XmnI (1) and the 1.3kb LCR was cloned into pBluescript, which had be cut with SpeI, blunted and then cut with BamHI, to generate the 1.3 LCR-pBS vector (2). The hCD2 LCR sequences were liberated from the 1.3 LCR-pBS vector by digestion with BamHI/XbaI (3). The 5.5kb LCR was deleted from the hCD2-Int4 vector by digestion with BamHI/XbaI (4) and replaced with the 1.3kb LCR generated in step 3 to give the hCD2-Int4-1.3 vector (5). The hCD2-Int4-1.3 vector was digested with XbaI/NotI and a single loxP site (represented as a black triangle) generated from annealed oligos cloned in (6).

Open boxes represent exons. K=KpnI. B=BamHI. Xm=XmnI. Xb=XbaI. N=NotI. S*=SpeI after blunting.

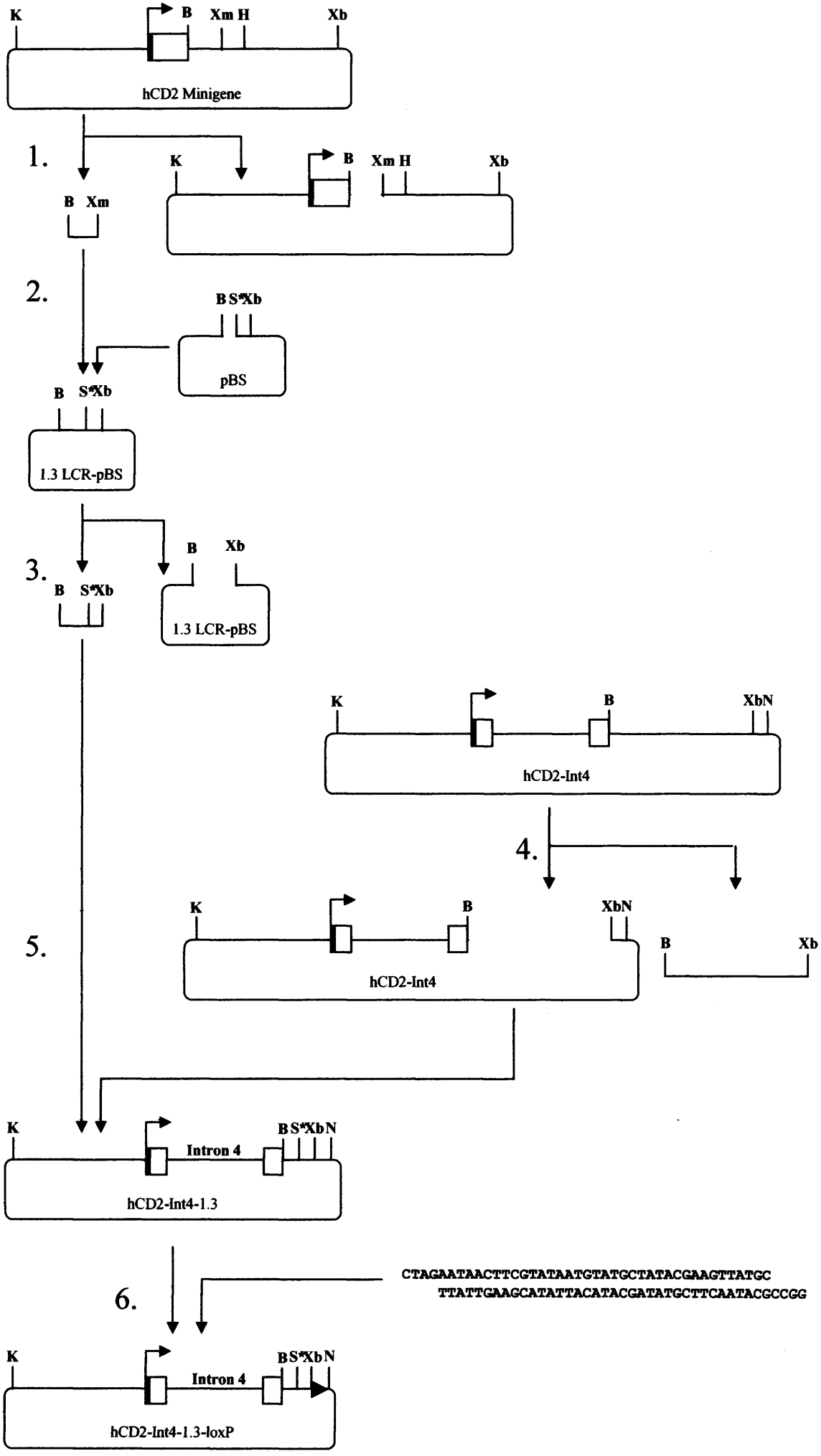

K
B
Xm H
Xb
hCD2 Minigene
1.
B Xm
K
B
Xm H
Xb
2.
B S*Xb
pBS
B S*Xb
1.3 LCR-pBS
3.
B
Xb
B S*Xb
1.3 LCR-pBS
K
B
XbN
hCD2-Int4
4.
5.
K
B
XbN
B
Xb
hCD2-Int4
K
Intron 4
B S*Xb N
hCD2-Int4-1.3
6.
CTAGAATAACTTCGTATAATGTATGCTATACGAAGTTATGC
TTATTGAAGCATATTACATACGATATGCTTCAATACGCCGG
K
Intron 4
B S*Xb N
hCD2-Int4-1.3-loxP

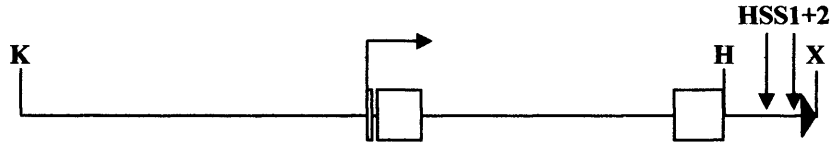
Figure 3. hCD2-Int4 expression constructs.

(A) The original hCD2-Int4 construct containing 5.5kb of downstream sequences. Hypersensitive sites (HSS) 1, 2 and 3 are indicated by arrows. Open boxes represent exons. K= KpnI. H=HinDIII. X=XbaI

(B) The hCD2-Int4-1.3loxP construct containing the truncated 1.3kb LCR is also shown. Hypersensitive sites 1 and 2 are indicated by arrows. Open boxes represent exons and filed triangle represents the loxP site. K= KpnI. H=HinDIII. X=XbaI

A. hCD2-Int4

B. hCD2-Int4-1.3-loxP



1KB

flow cytometry (data not shown). Seven founder mice were bred to CBA non-transgenic mice to establish seven independent transgenic lines denoted 1.3loxP.1-7.

3.3 Analysis of 1.3loxP transgenic mice

3.3.1 Flow cytometric analysis of peripheral T cells from 1.3loxP transgenic mice

To determine whether inclusion of intron 4 could compensate for the absence of HSS3 in transgenic mice (i.e. prevent variegation) transgene expression profiles on peripheral T cells in 1.3loxP transgenic mice were characterised by flow cytometry. Single cell suspensions from spleen were triple stained with anti-CD4, anti-CD8 and anti-hCD2 before analysis by flow cytometry. Flow cytometry data from 1.3loxP.1 and 1.3loxP.4 lines are shown as examples (figure. 4A). Splenocytes from an MG4 transgenic mouse were used as a positive control. The MG4 (homozygous) line contains ~12 copies of the hCD2 minigene with the full LCR, and does not exhibit position effect variegation even though the transgenes are integrated proximal to a centromere [342]. Analysis of MG4 $CD4^+$ splenocytes revealed a unimodal flow cytometry profile that is characteristic of LCR directed expression with more than 99% of $CD4^+$ splenocytes expressing high levels of hCD2 (mean fluorescent intensity [MFI] of 790). In contrast to the MG4, 1.3loxP.1 $CD4^+$ splenocytes exhibited two populations that either expressed hCD2 at high levels (74%/MFI 530) or expressed hCD2 at low levels or not at all (26%). The finding that not all cells express hCD2 indicates that the hCD2 transgene in the 1.3loxP.1 line is subject to PEV. Analysis of 1.3loxP.4 $CD4^+$ splenocytes revealed an expression profile more similar to that of the MG4. Virtually all of the 1.3loxB $CD4^+$ splenocytes expressed high levels of hCD2 (MFI 223), indicating that the transgene in this line is not subject to PEV at this integration site.

3.3.2 Slot blot analysis of 1.3loxP.1 and 1.3loxP.4 transgenic lines

Genomic tail DNA was isolated from 1.3loxP.1 and 1.3loxP.4 transgenic mice and subjected to slot blot analysis to determine the transgene copy number of each line. Slot blots were probed with the hCD2 cDNA and with the mCD2 cDNA to allow for loading normalisation. Transgenic lines of known hCD2 copy number were included as controls. Comparison to the lines of known copy number allows the approximate transgene copy number to be determined (note that as only for the hCD2.3 line has the exact copy number been determined all calculated copy numbers are approximate). Figure 4B shows the calculated transgene copy number of each line. The 1.3loxP.1 line is calculated to contain approximately 6 copies of the hCD2 transgene, whereas the 1.3loxP.4 is calculated to contain approximately 1-2 copies of the hCD2 transgene.

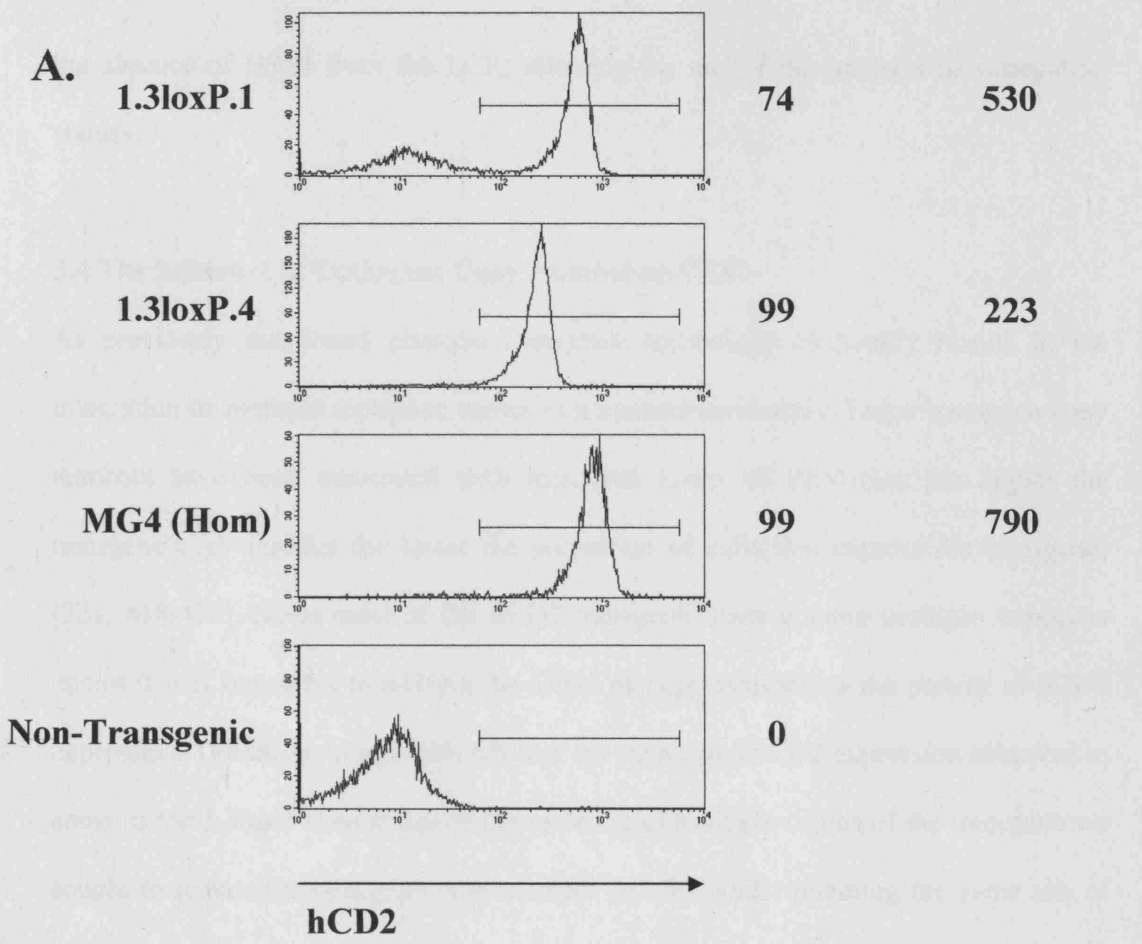
Comparing transgene copy number analysis with flow cytometry data allows the relative level of expression per transgene copy in each line to be calculated. For example, the 1.3loxP.4 line contains ~1-2 transgene copies, but expresses high levels of hCD2 (MFI 223 on $CD4^+$ splenocytes). In comparison the MG4 line contains ~12 copies of the hCD2 minigene (that does not carry intron 4,) but express less than 4 times the level of hCD2 (MFI 790 on $CD4^+$ splenocytes) as 1.3loxP.4 transgenic mice. This confirms that inclusion of intron 4 results in increased levels of expression per transgene copy.

In summary, inclusion of intron 4 results in increased levels of expression when compared to the hCD2 mini gene, as previously noted, thus, potentially allowing the examination of single copy transgenic mice. Importantly, however, intron 4 cannot compensate for

Figure 4. hCD2 expression and transgene copy number in 1.3loxP.1 and 1.3loxP.4 transgenic mice.

(A) Splenocytes were stained with antibodies specific for CD4, CD8 and hCD2. Histograms show expression of hCD2 on CD4 single positive lymphocytes. Values are the percentage of each single positive population expressing hCD2 and the mean fluorescence intensity of anti-hCD2-FITC on the gated population, respectively.

(B) Samples containing 10μg, 5μg, 2.5μg of genomic DNA were each equally loaded into two separate wells. Slots were probed with either hCD2 cDNA or mCD2 to allow normalisation of loading. Copy number was determined by comparison to the following transgenic lines CD2.3 (homozygous), MG4 (homozygous), CD2.2B (homozygous) which contain 2, 12, and 32 copies of the hCD2 gene, respectively.


%hCD2+
MFI on hCD2+ cells
A.
1.3loxP.1
74
530
1.3loxP.4
99
223
MG4 (Hom)
99
790
Non-Transgenic
0
hCD2
B.
Transgenic Line
mCD2
hCD2
Copy Number
1.3loxP.1
6
1.3loxP.4
1-2
B10
0
MG4 (HOM)
12
CD2.2B
32
CD2.3
2
10
5
2.5

the absence of HSS3 from the LCR, allowing the use of the intron 4 in variegation studies.

3.4 The Influence of Transgene Copy Number on PEV

As previously mentioned classical transgenic technology frequently results in the integration of multiple transgene copies in a concatameric array. Large transgene copy numbers have been associated with increased levels of PEV (i.e. the higher the transgene copy number the lower the proportion of cells that express the transgene) [328, 418-423]. Since most of the hCD2 transgenic lines contain multiple transgene copies it was important to address the affect of copy number on the pattern of hCD2 expression. Therefore, to establish whether the variegated hCD2 expression observed in some os the 1.3loxP lines is due to the presence of multiple copies of the transgene we sought to reduce the transgene copy number *in vivo* whilst retaining the same site of integration. As mentioned above a single loxP site was included at the 3' of the 1.3loxP construct. Integration of the multiple transgene copies in the 1.3loxP lines in a head to tail orientation would theoretically allow Cre mediated recombination between loxP sites on adjacent copies of the transgene and the excision of the multiple copies leaving behind a single transgene copy. To mediate this transgene copy number reduction each 1.3loxP line was bred to the protamine-Cre (PC3-Cre) transgenic mouse line, in which Cre expression is driven by the protamine 1 gene regulatory elements [424]. As the expression of the PC3-Cre transgene is restricted to the haploid spermatids, deletion of the floxed allele occurs only on transgene transmission through the male germline [424].

From the progeny of 1.3loxP/PC3 double transgenic males, mice that displayed an altered hCD2 expression profile (data not shown) were selected for further breeding to establish new lines denoted by the suffix 'Del' in which the PC3-Cre transgene had been bred out.

3.4.1 Slot Blot Analysis of 1.3loxP and 1.3loxPDel Transgenic Lines

Genomic tail DNA was isolated from 1.3loxP (referred to as the parental lines) and 1.3loxPDel transgenic mice and subjected to slot blot analysis to determine the approximate transgene copy number of each line. It is important to note that this copy number analysis does not give exact copy numbers, but rather allows the changes in copy number to be analysed. Slot blots were probed with the hCD2 cDNA and with the mCD2 cDNA to allow for loading normalisation. Figure 5 shows the calculated approximate transgene copy number in each of the 1.3loxPDel transgenic lines as compared to the parental 1.3loxP lines. Transgenic lines of known hCD2 copy number were included as controls. Comparison to lines of known copy number allows the approximate transgene copy number to be determined (note that only for the hCD2.3 line has the exact copy number been determined). Copy number analysis revealed that in all but one of the 1.3loxPDel transgenic lines the copy number had been decreased through Cre mediated recombination (see figure for calculated copy numbers). In the 1.3loxP.1 line the copy number had been decreased from ~6 copies to ~4 copies of the transgene to generate the 1.3loxP.1Del line. Similarly, in the 1.3loxP.3 line the copy number had been decreased from ~10 copies to approximately ~3 copies to generate the 1.3loxP.3Del line. In the 1.3loxP.6 line the copy number had been decreased from ~20 copies to ~6 copies of the transgene to generate the 1.3loxP.6Del line. Finally, in the 1.3loxP.7 line the copy number had been decreased from ~6 copies to ~4 copies of the

transgene to generate the 1.3loxP.7Del line. In contrast the 1.3loxP.5 parental transgenic line was calculated to contain ~1-2. No decrease in transgene copy number was detected in the 1.3loxP.5Del transgenic line.

These results reveal that the crossing the 1.3loxP transgenic lines to the PC3-Cre transgenic line provides an effective system to reduce transgene copy numbers whilst retaining the same site of integration. Interestingly, however, in none of the transgenic lines had the transgene array been reduced to a single copy (see below for Southern blot analysis). This suggests that, for a reason yet to be undetermined, Cre mediated recombination had not reached completion in each of the transgenic lines.

3.4.2 Southern Analysis of 1.3loxP and 1.3loxPDel Transgenic Lines

Southern blot analysis was used to analyse the transgene structure in each of the transgenic lines (note that as loading controls were not performed Southern blot analysis could not used for copy number analysis). Figure 6 shows Southern blot analysis of the 1.3loxP and 1.3loxPDel transgenic lines. DNA from 1.3loxP and 1.3loxPDel transgenic mice was digested with BamHI and subjected to Southern blot analysis. hCD2 cDNA was used as a probe (figure 6B). Genomic DNA from a B10 non-transgenic mouse and an MG4 (homozygous) mouse was included as a negative and positive control, respectively. In the lane corresponding to the MG4 DNA a band of approximately 12kb is evident. This corresponds to ~12 copies of the hCD2 transgene integrated in a head to tail orientation.

In the lane corresponding to the 1.3loxP.1 line three bands are evident. The 12kb band corresponds to the expected size that would be produced if multiple copies of the

transgene were integrated in a head to tail orientation (see figure 6C). The band of approximately 20kb corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to head orientation (see figure 6C). The largest band most likely represents an end fragment. Thus, in the 1.3loxP.1 line transgene copies have integrated in both the head to tail and head to head orientation. No major transgene rearrangements are obvious. Reducing the transgene copy number to produce the 1.3loxP.1Del line did not result in any significant alteration in the Southern profile.

In the lane corresponding to the 1.3loxP.2 line three bands are also evident. The band of approximately 20kb most likely represents multiple copies of the transgene integrated in a head to head orientation. In contrast the two bands of 9kb and 3kb are too small to have originated from complete copies of the transgene even as an end fragment (which would have a minimum possible size of ~10kb). The fragments most probably originate from incomplete end fragments or rearranged transgene copies. Due to the evidence for transgene rearrangements this line was not chosen for further analysis.

In the lane corresponding to the 1.3loxP.3 line three bands are evident. The 12kb band corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to tail orientation. The band at approximately 20kb corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to head orientation. The largest band most likely represents an end fragment. Thus, in the 1.3loxP.3 line transgene copies have integrated in both the head to tail and head to head orientation. In addition, there appears to be more copies of the transgene integrated in the head to head orientation than in the head

Figure 5. Slot blot analysis of transgene copy number in 1.3loxP and 1.3loxP.Del transgenic mice.

Samples containing 10μg, 5μg, and 2.5μg of genomic DNA were each equally loaded into two separate wells. Slots were probed with either hCD2 cDNA or mCD2 to allow normalisation of loading. Copy number was determined by comparison to the following transgenic lines CD2.3 (homozygous), MG4 (homozygous), CD2.2B (homozygous) which contain 2, ~12, and ~32 copies of the hCD2 gene, respectively. Calculated copy numbers are indicated to the right of the slot blot.

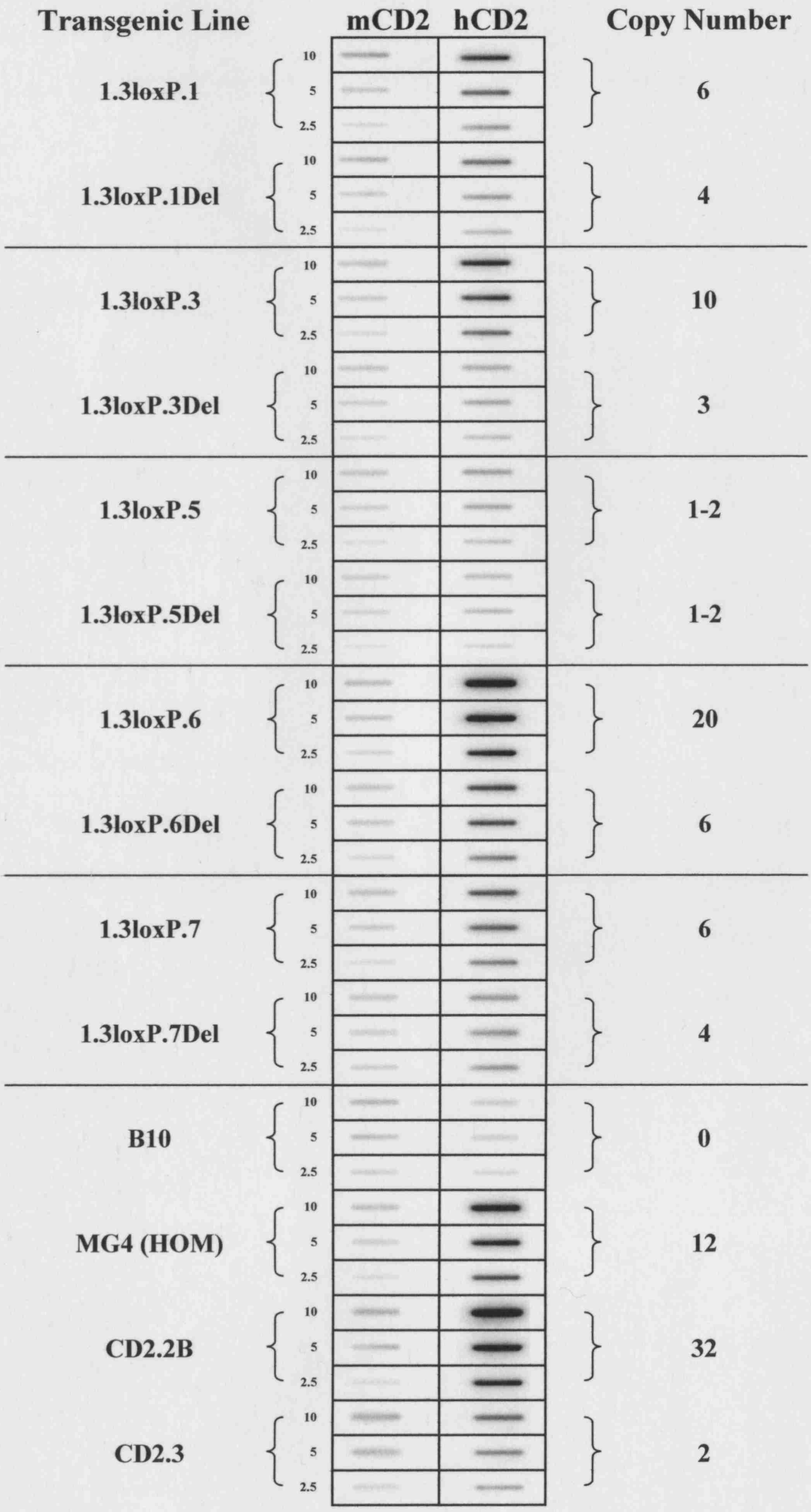

Transgenic Line
mCD2
hCD2
Copy Number
1.3loxP.1
10
5
2.5
6
1.3loxP.1Del
10
5
2.5
4
1.3loxP.3
10
5
2.5
10
1.3loxP.3Del
10
5
2.5
3
1.3loxP.5
10
5
2.5
1-2
1.3loxP.5Del
10
5
2.5
1-2
1.3loxP.6
10
5
2.5
20
1.3loxP.6Del
10
5
2.5
6
1.3loxP.7
10
5
2.5
6
1.3loxP.7Del
10
5
2.5
4
B10
10
5
2.5
0
MG4 (HOM)
10
5
2.5
12
CD2.2B
10
5
2.5
32
CD2.3
10
5
2.5
2

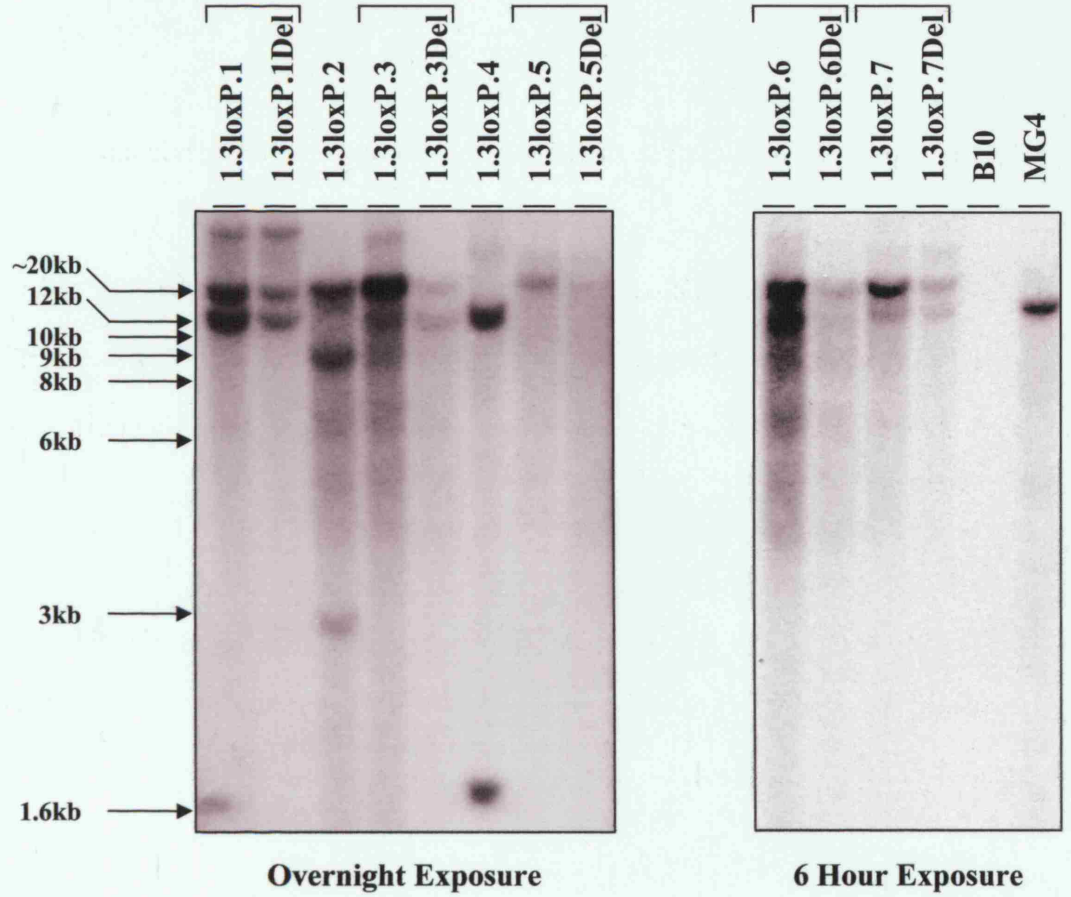
Figure 6. Southern blot analysis of 1.3loxP and 1.3loxP.Del transgenic lines.

(A) Genomic DNA from 1.3loxP and 1.3loxP.Del transgenic mice was digested with BamHI. Genomic DNA from a B10 non-transgenic mouse included as a negative control. Genomic DNA from an MG4 mouse was included as a positive control. The probe is indicated below in (B).

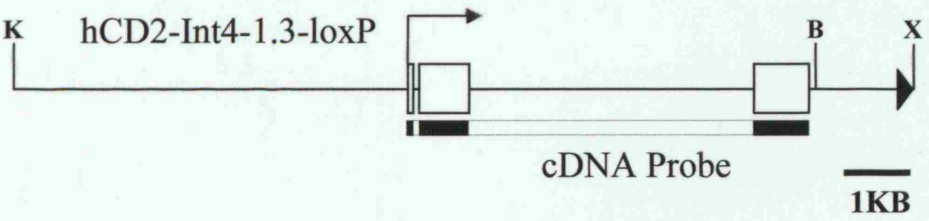
(B) Schematic diagram of the 1.3loxP construct which was used to generate the ΔMAR transgenic lines. Positions of relevant restriction sites and cDNA probe are indicated. Open boxes represent exons. B=BamHI, K=KpnI, X=XbaI.

(C) Schematic diagram of the expected BamHI fragments from 1.3loxP transgene copies integrated in a head to tail, head to head and tail to tail orientation. The red solid line represents the fragments that would be detected with the cDNA probe. The red dotted line represents the BamHI fragments that would not be detected with the cDNA probe. Open boxes represent exons. B=BamHI.

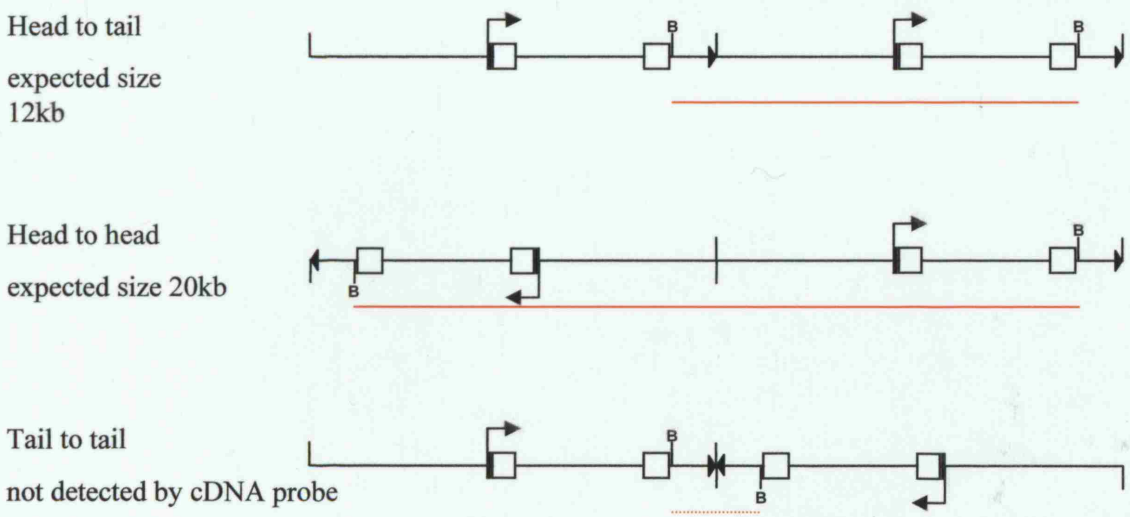
A.


1.3loxP.1
1.3loxP.1Del
1.3loxP.2
1.3loxP.3
1.3loxP.3Del
1.3loxP.4
1.3loxP.5
1.3loxP.5Del
1.3loxP.6
1.3loxP.6Del
1.3loxP.7
1.3loxP.7Del
B10
MG4
~20kb
12kb
10kb
9kb
8kb
6kb
3kb
1.6kb
Overnight Exposure
6 Hour Exposure

B.


K
hCD2-Int4-1.3-loxP
B
X
cDNA Probe
1KB

C.


Head to tail
expected size
12kb
Head to head
expected size 20kb
Tail to tail
not detected by cDNA probe
B

to tail orientation. No major transgene rearrangements are obvious. Reducing the transgene copy number to produce the 1.3loxP.3Del line did not result in any significant alteration in the Southern profile.

In the lane corresponding to the 1.3loxP.4 line three bands are evident. The 12kb band most likely represents multiple copies of the transgene integrated in a head to tail orientation. The 1.6kb band is too small to have originated from a complete copy of the transgene and the signal is too intense to have originated from a single copy of the transgene. This band therefore most likely originated from rearranged copies of the transgene. The largest band most likely represents an end fragment. Slot blot analysis reveals that the 1.3loxP.4 line contains ~1-2 copies of the transgene. Southern analysis reveals that there are most probably at least 2 complete copies of the transgene in this line. Due to the evidence for rearrangements this lines was not chosen for further analysis.

In the lane corresponding to the 1.3loxP5 line two bands are evident. The band of approximately 20kb corresponds to the expected size of two copies of the transgene integrated in a head to head orientation. Alternatively, this band may also represent an end fragment. The presence of two bands in this lane rules out the possibility that this line contains only a single transgene copy. Importantly, both bands have been noted in previous Southern blots. The lane corresponding to the 1.3loxP.5Del line reveals a similar Southern profile, although the signal intensity of the 20kb band is decreased suggesting that there are fewer copies of the transgene in the 1.3loxP.5Del line. However, without loading controls it is not possible to quantify this decrease, if any. As this 20kb band corresponds to the expected size of two copies of the transgene

integrated in a head to head orientation if there was a copy number reduction to produce the 1.3loxP.5Del line the parental line would have to contain atleast 4 copies of the transgene. This is clearly not in agreement with the copy number calculated by slot blot analysis. Indeed, when compared to the CD2.3 line (note that the exact copy number of the CD2.3 line has been determined as one copy in the heterozygous state) the 1.3loxP.5 line is calculate to contain ~1-2 copies of the transgene. Thus, the Southern blot will need to be repeated with loading controls and additional restriction enzymes.

In the lane corresponding to the 1.3loxP.6 line three bands are evident. The 12kb band corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to tail orientation. The band of approximately 20kb corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to head orientation. The largest band, which is very faint due to the short exposure time, most likely represents an end fragment. Thus, in the 1.3loxP.6 line transgene copies have integrated in both the head to tail and head to head orientation. No major transgene rearrangements are obvious. Reducing the transgene copy number to produce the 1.3loxP.6Del line did not result in any significant alteration in the Southern profile.

In the lane corresponding to the 1.3loxP.7 line three bands are evident. The 12kb band corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to tail orientation. The band of approximately 20kb corresponds to the expected size that would be produced if multiple copies of the transgene were integrated in a head to head orientation. The largest band, which is faint due to the short exposure time, most likely represents an end fragment. Thus, in the

1.3loxP.7 line the majority of the transgene copies have integrated in the head to head orientation and a small proportion in the head to tail orientation. No major transgene rearrangements are obvious. Interestingly, reducing the transgene copy number to produce the 1.3loxP.7Del line did not result in any significant alteration in the Southern profile.

3.4.3 Flow Cytometric Analysis of Peripheral T Cells after Transgene Copy Number Reduction

To characterise the affect of reducing the transgene copy number on hCD2 expression in 1.3loxP transgenic mice, flow cytometry was performed on purified peripheral T cells from 1.3loxP and 1.3loxPDel transgenic mice. Single cell suspensions from spleen were triple stained with anti-CD4, anti-CD8 and with anti-hCD2 before analysis by flow cytometry. Figure 7 shows a representative example from each line. Figure 8 shows statistical data obtained from analysis of multiple mice from each transgenic line.

Comparison of hCD2 expression on 1.3loxP and 1.3loxPDel $CD4^+$ T cells reveals that in each 1.3loxDel line the reduction in copy number is associated with a paralleled decreased in the level of expression (as indicated by the MFI). However, the decrease in the level of expression is not directly proportional to the decrease in transgene copy number. For example, in the 1.3loxP.3 line the 3 fold decrease in transgene copy number to generate the 1.3loxP.3Del line (the copy number was reduced from ~10 copies to ~3 copies) is only accompanied by a 2 fold decrease in the level of expression. In addition to a decrease in the level of expression, reducing the transgene copy number affected the pattern of variegation in a number of lines. For example, in the 1.3loxP.1 line the percentage of cells expressing hCD2 is ~70%. However, in the 1.3loxP.1Del

Figure 7. Expression of hCD2 on CD4 single positive splenocytes isolated from 1.3loxP and 1.3loxP.Del transgenic mice.

Splenocytes were stained with antibodies specific for CD4, CD8 and hCD2. Histograms show expression of hCD2 on CD4 single positive lymphocytes. Values are the percentage of each single positive population expressing hCD2 and the mean fluorescence intensity of anti-hCD2-FITC on the gated population, respectively.

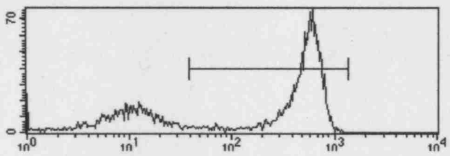
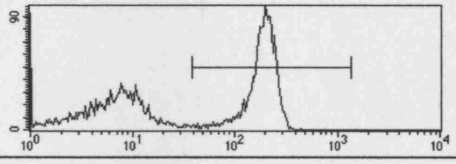
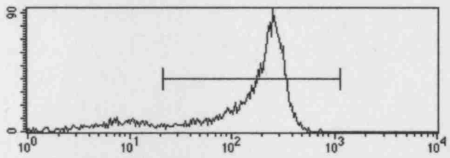
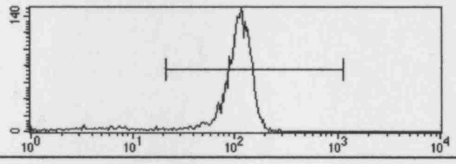
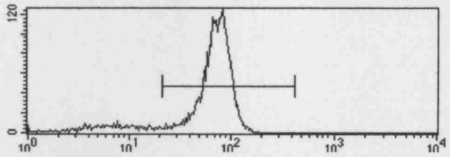
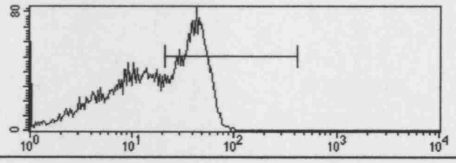
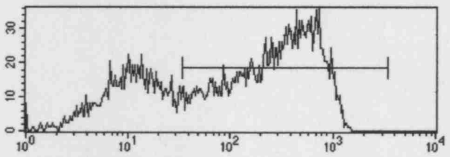
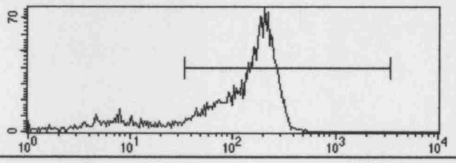
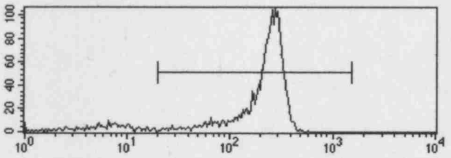
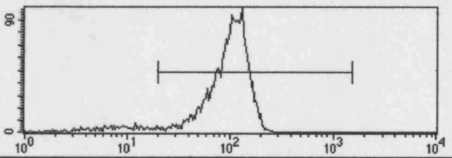
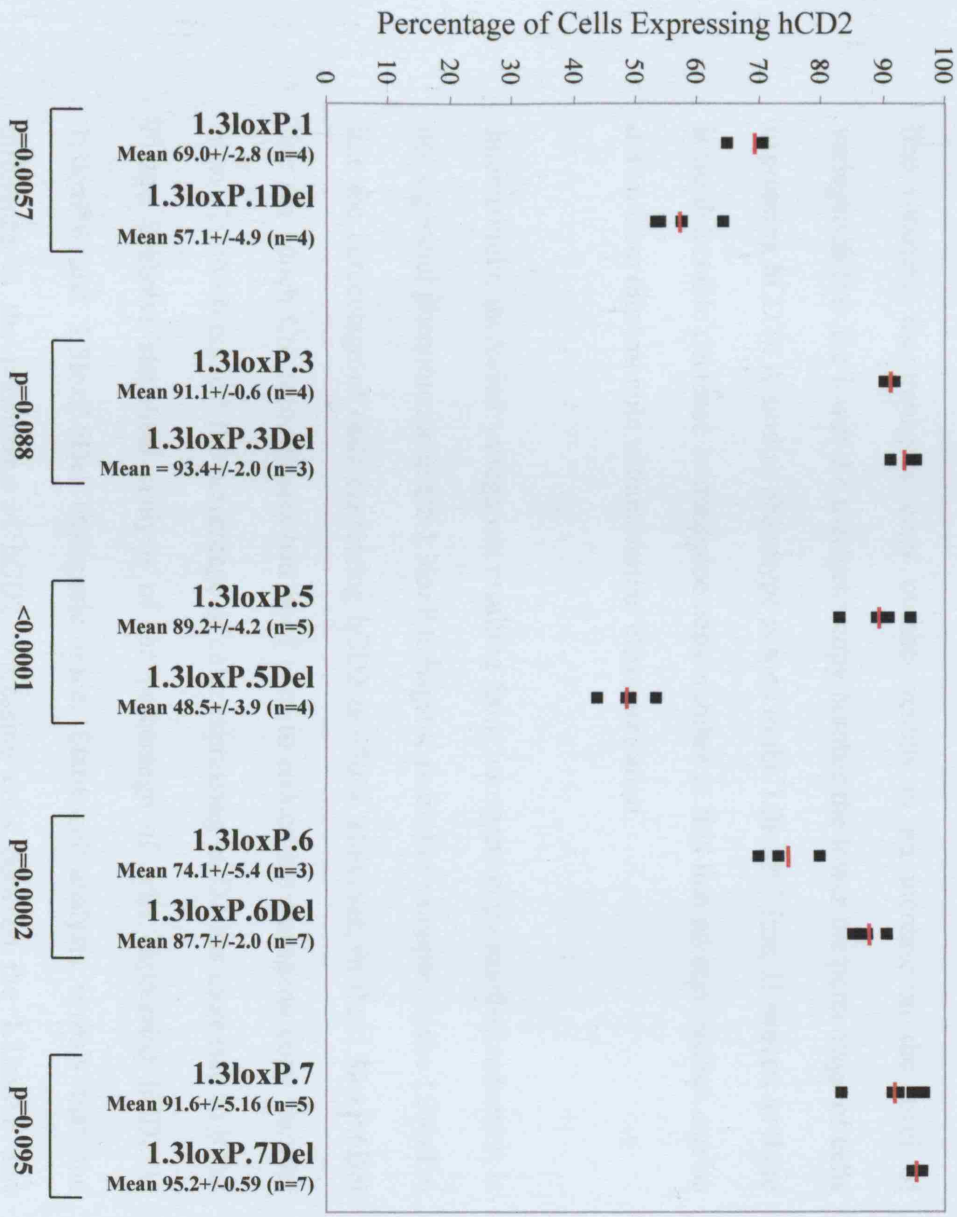
		%hCD2$^+$	MFI on hCD2$^+$ cells
1.3loxP.1		70%	508
1.3loxP.1Del		57%	184
1.3loxP.3		90%	208
1.3loxP.3Del		95%	108
1.3loxP.5		91%	72
1.3loxP.5Del		49%	40
1.3loxP.6		70%	372
1.3loxP.6Del		87%	163
1.3loxP.7		92%	227
1.3loxP.7Del		95%	107

Figure 8. Statistical analysis of the percentage of hCD2 expressing cells in 1.3loxP and 1.3loxP.Del transgenic mice.

Graph showing the percentage of hCD2 expressing cells in 1.3loxP and 1.3loxP.Del transgenic mice. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated. N= the number of mice in each data set.


Percentage of Cells Expressing hCD2
0
10
20
30
40
50
60
70
80
90
100
1.3loxP.1
Mean 69.0+/-2.8 (n=4)
1.3loxP.1Del
Mean 57.1+/-4.9 (n=4)
p=0.0057
1.3loxP.3
Mean 91.1+/-0.6 (n=4)
1.3loxP.3Del
Mean = 93.4+/-2.0 (n=3)
p=0.088
1.3loxP.5
Mean 89.2+/-4.2 (n=5)
1.3loxP.5Del
Mean 48.5+/-3.9 (n=4)
<0.0001
1.3loxP.6
Mean 74.1+/-5.4 (n=3)
1.3loxP.6Del
Mean 87.7+/-2.0 (n=7)
p=0.0002
1.3loxP.7
Mean 91.6+/-5.16 (n=5)
1.3loxP.7Del
Mean 95.2+/-0.59 (n=7)
p=0.095

line (in which Cre recombinase had been used to reduce the transgene copy number from ~6 to ~4 copies) the percentage of cells expressing hCD2 has decreased to 57%. Figure 8 shows statistical analysis of the percentage of cells expressing hCD2 in 1.3loxP.1 and 1.3loxP.1Del transgenic mice. Statistical analysis reveals that the difference in the percentage of hCD2 expressing cells between the 1.3loxP.1 and 1.3loxP.1Del is statistically significant (p=0.0057). This suggests that in the 1.3loxP.1 line reducing the transgene copy number results in an increase in the level of variegation (i.e. the lower the transgene copy number the lower the percentage of cells expressing hCD2). A similar phenotype is seen in the 1.3loxP.5 line. However, as there is no detectable decrease in transgene copy number in this line no conclusions can be drawn from this example without further characterisation.

Interestingly, increased variegation resulting from transgene copy number reduction is not a general phenomenon in the 1.3loxP transgenic lines. For example, in the 1.3loxP.6 line the percentage of cells expressing hCD2 is ~70%. However, in the 1.3loxP.6Del line (in which Cre recombinase had been used to reduce the transgene copy number from ~20 to ~6 copies) the percentage of cells expressing hCD2 has increased to 87%. Figure 8 shows statistical analysis of the percentage of cells expressing hCD2 in 1.3loxP.6 and 1.3loxP.6Del transgenic mice. Statistical analysis reveals that the difference in the percentage of hCD2 expressing cells between the 1.3loxP.6 and 1.3loxP.6Del is statistically significant (p=0.0002). This suggests that in the 1.3lox.6 line reducing the transgene copy number results in a decrease in the level of variegation (i.e. the lower the transgene copy number the greater the percentage of hCD2 expressing cells).

In contrast, reducing the transgene copy number in the 1.3loxP.3 and 1.3loxP.7 lines did not result in a significant change in the level of variegation (figure 7). Figure 8 shows that there in no statistical significance in the percentage of hCD2 expressing cells resulting from copy number reduction in the 1.3loxP.3 and 1.3loxP.7 transgenic lines (P=0.088 and 0.095 respectively).

In summary, Cre mediated recombination is an efficient system to reduce transgene copy number *in vivo*, whilst retaining the same integration site. However, in none of the transgenic lines was the transgene array reduced to a single copy (as shown by Southern blot analysis). Flow cytometry revealed that deleting transgene copies from variegating 1.3loxP transgenic lines resulted in a decrease in the level of transgene expression. Interestingly, the decreases in transgene expression levels were not directly proportional to the decreases in transgene copy number. In addition to affecting levels of expression, deleting transgene copies resulted in either an increase in the level of variegation (1.3loxP.1), a decrease in the level variegation (1.3loxP.6) or no change in the level of variegation (1.3loxP.3 and 1.3loxP.7). Thus, it appears that deleting copies of the 1.3loxP transgene *in vivo* whilst retaining the same integration site affected the pattern of variegation in a manner that was specific to each transgenic line. Therefore, the relathionship between variegation and copy number is dependant upon the site of integration in the hCD2 system.

3.5 Imprinting in the 1.3lox Lines

Classical genomic imprinting is a form of gene regulation that results in a specific subset of mammalian genes being expressed from only one of the two parental chromosomes [425]. Some imprinted genes are expressed from the maternally inherited

chromosome while others are expressed from the paternally inherited chromosome [425]. Such imprinting is an important mechanism for the regulation a number of genes, such as *IGF2* and *IGF2r* [425].

Within the germline genes can be specifically 'marked' or 'imprinted' in a manner that determines their pattern of expression in the progeny [425]. The principle mechanism utilised to imprint genes is DNA methylation, although histone modifications and modulation of chromatin structure have also been implicated [425]. Within the primordial germ cells the entire genome is actively demethylated removing almost all of the parental imprints. During gametogenesis *de novo* methylation occurs establishing new germ line specific methylation imprints [425]. Thus, specific genes are marked differently in the male and female germ lines. These specific methylation marks are retained after fertilisation and throughout development, thereby resulting in the differential regulation of paternally and maternally inherited genes [425]. The mechanisms by which methylation marks are placed differently in males and females are not fully understood [425].

Unexpectedly, whilst backcrossing the 1.3loxP lines to the CBA background it became evident that the 1.3loxP hCD2 transgene was subject to a form of imprinting in a number of transgenic lines. In 4 out of the 7 lines generated, the pattern of hCD2 expression was seen to be dependent on the parental origin of the transgene.

Figure 9 shows the comparison of hCD2 variegation patterns in mice with paternally or maternally derived 1.3loxP transgenes. Single cell suspensions from spleen were triple stained with anti-CD4, anti-CD8 and with anti-hCD2 before analysis by flow

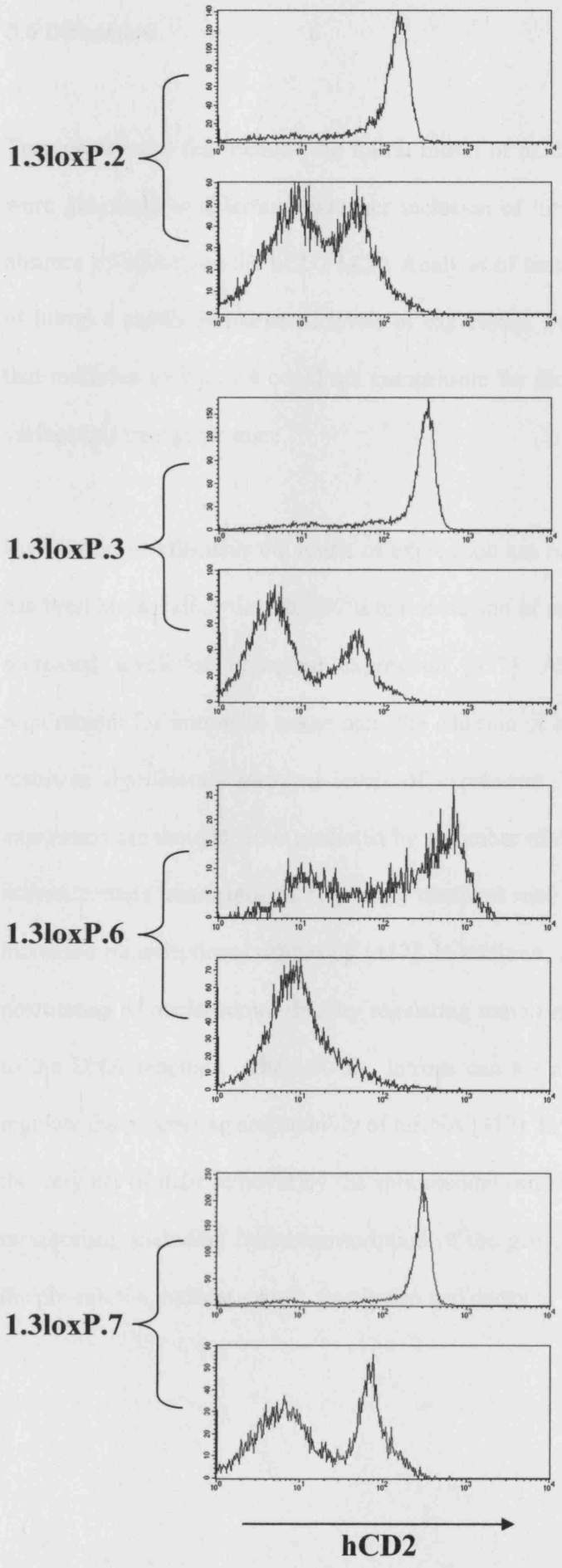
cytometry. Within each line comparison of the levels of hCD2 expression from transgenes of paternal or maternal origin reveals that in all cases the level of hCD2 expression is dramatically lower when the transgene is of a maternal origin. In addition, it appears that the maternally derived transgenes are more susceptible to PEV. Interestingly, the pattern of hCD2 expression is indistinguishable in male and female littermates (data not shown) indicating that the variegation profile is not dependent upon the sex of the mouse being analysed, but the parental origin of the transgene. Furthermore, although male 1.3loxP transgenic mice that receive a maternally derived transgene express low levels of hCD2, their offspring express high levels of hCD2, suggesting that the imprinting event can be reversed or erased on passage through the male germline (data not shown). The mechanisms responsible for this imprinting have not been characterised. Imprinting in the additional three 1.3loxP lines has yet to be investigated. However, in the 1.3loxP.1 line the transgene is integrated on the Y chromosome (as determined by breeding analysis) and as a result female mice are unable to carry the transgene preventing imprinting analysis in this line.

Due to imprinting within the 1.3loxP lines sections 3.3-3.4 contain only analysis of 1.3loxP transgenic mice in which the transgene was of the paternal origin.

Figure 9. Imprinted hCD2 expression in 1.3loxP transgenic mice.

Expression of hCD2 on CD4 single positive splenocytes isolated from 1.3loxP transgenic mice. Splenocytes were stained with antibodies specific for CD4, CD8 and hCD2. Histograms show expression of hCD2 on CD4 single positive lymphocytes from mice in which the hCD2 transgene was of either maternal or paternal origin. The sex of the parent from which the transgene originated is indicated to the right of the histogram (note that this label does not refer to the sex of the mouse used in the analysis).

Parental Origin of Transgene



3.6 Discussion

Transgenic mice that include the fourth intron of hCD2 and the truncated 1.3kb LCR were generated to determine whether inclusion of intron 4 could compensate for the absence of HSS from the hCD2 LCR. Analysis of these mice confirmed that inclusion of intron 4 results in increased levels of expression. Furthermore, it was demonstrated that inclusion of intron 4 could not compensate for the loss of HSS3 from the LCR in variegating transgenic mice.

How intron 4 influences the levels of expression has not been determined. However, it has been known since the late 1970s that inclusion of intronic sequences often results in increased levels of transgene expression [417]. Although there is no universal requirement for introns in eukaryotes, the addition of just one generic intron can often result in significantly elevated levels of expression [417]. These elevated levels of expression are thought to be mediated by a number of mechanisms. For example, many introns contain transcriptional regulatory elements such as enhancers, which may confer increased transcriptional efficiency [417]. In addition, introns are able to modulate the positioning of nucleosomes thereby regulating transcription by controlling accessibility to the DNA template. Alternatively, introns can act at a post-transcriptional level to regulate the processing and stability of mRNA [417]. In fact, the presence of introns and the very act of their removal by the spliceosome can influence many stages of mRNA metabolism, including initial transcription of the gene, editing and polyadenylation of the pre-mRNA, nuclear export, translation and decay of the mRNA [417].

In transient transfection assays the HSS of intron 4 is unable to function as a classical enhancer. It is therefore more likely that the elevated levels of expression from the hCD2-Int4-1.3-loxP cassette result from increased mRNA stability imparted by the inclusion additional intronic sequences.

3.6.1 Transgene Copy Number and Variegation

It has been proposed that PEV in transgenic mice and also *Drosophila* is as a direct result of the repetitive nature of transgene arrays [328, 418-423]. This phenomenon has been termed 'repeat-induced silencing'. In addition, nucleotide triplet-repeat expansion has also been associated with increased levels of transgene PEV [426]. Interestingly, heterochromatin rich regions such as centromeres and telomeres are mainly composed of repetitive sequences and are able to induce transgene silencing and PEV [341]. It is thought that these repetitive sequences allow more regular folding of the DNA and therefore promote the formation of heterochromatin. However, the formation of heterochromatin is likely to be dependent on the sequence composition of the repeat; for example a number of endogenous genes naturally exist as concatameric arrays (such as the rRNA, tRNA and histone genes), which are not packaged into heterochromatin [418].

To directly address the influence of transgene copy number on PEV, a single loxP site was included at the 3' end of the 1.3loxP transgene. Transgenic Cre activity in multicopy 1.3loxP transgenic mice resulted in a reduction of copy number in each line (except in the 1.3loxP.5 parental line which was calculated to contain ~1-2 copies of the transgene). Interestingly, in none of the lines had Cre mediated recombination reached completion by reducing the transgene array to a single copy (as shown by Southern blot

analysis). The reason for incomplete recombination in these lines has not been determined. However, there are a number of potential explanations. First, the loxP site is at the very 3' end of the 1.3loxP construct. On integration into the genome it is possible that the loxP site may have been lost from a number of copies of the transgene. If these copies where not flanked by loxP containing copies of the transgene then they could not be deleted upon Cre mediated recombination. Alternatively, if transgene copies were integrated in opposing orientations recombination between loxP sites in opposing directions would result in inversion of the intervening sequences and not deletion. Indeed, most of the 1.3loxP lines contain transgene copies integrated in opposing orientations. However, Cre mediated recombination within an array containing transgene copies in opposing orientations would always be resolved to a maximum of two transgene copies, providing the recombination reaction reached completion.

Finally it is important to note that Cre mediated recombination in PC3-Cre transgenic mice is not 100% efficient. For example, Walmsley et al. reported the generation of *Rac1* knockout mice [427]. During targeting to the *Rac1* locus three loxP sites were introduced into the genome. On passage of the targeted allele through the male germline of PC3-Cre transgenic mice, Cre mediated recombination was observed at the targeted *Rac1* locus. Interestingly, all possible recombination events were noted in the resulting offspring, indicating that PC3-Cre directed recombination was not 100% efficient (Walmsley et al. supporting online material). Thus the failure to reduce the transgene array to a single copy in the majority of 1.3loxPDel transgenic lines may also reflect an inherent inefficiency of the PC3-Cre line.

In each of the 1.3loxPDel lines the decrease in copy number was associated with a concomitant decrease in the level of hCD2 surface expression on T cells. Interestingly, however, the decreases in expression levels were not directly proportional to the decrease in copy numbers. This can be explained by the fact that these transgenic lines contain a disabled hCD2 LCR and therefore true copy number-dependent expression is not conferred.

In addition to the decrease in the levels of expression, copy number reduction influenced the pattern of variegation in a number of lines. Interestingly, deleting transgene copies affected the pattern of variegation differently in different transgenic lines. As these lines differ in site of transgene integration (this is based on assumption and has not been formally proven) it is possible that the position within the genome influences the outcome of deleting transgene copies.

Previously, Garrick et al. reported a similar Cre based strategy to reduce copy number of an α globin-lacZ transgene, whilst retaining the same integration site [418]. Two high copy number α globin-lacZ transgenic mouse lines were generated. In both of these lines <1% of the expected cells expressed the transgene. Cre mediated recombination was subsequently used to reduce the transgene copy number to generate low copy number α globin-lacZ transgenic mice whilst retaining the same integration site. In the low copy number lines the number of cells expressing the transgene had increased to >50%. These data suggest that high transgene copy numbers induce gene silencing. These results are dissimilar to those reported herein that suggest that higher transgene copy numbers can in some instances help overcome PEV. Although the

mechanisms responsible for these differences have not been investigated there are a number or potential explanations.

First, the data reported herein utilised a mammalian promoter and a mammalian reporter gene, whilst the α globin-lacZ transgene relied upon the prokaryotic reporter gene lacZ. It must be noted that genes of prokaryotic origin are exceptionally susceptible to PEV [428]. Furthermore, as lacZ is so unusually CpG rich (lacZ has a CpG content of >9% compared to an average of 1% in the human genome) it is predisposed to methylation induced silencing and PEV [429]. Increasing the number of lacZ copies generates a larger area for CpG methylation, therefore increasing the probability of transgene silencing, predisposing lacZ transgenes to repeat-induced silencing. The observation that lacZ is able to render the complete β-globin LCR sensitive to PEV highlights the potential of this transgene to cause transcriptional deregulation [430, 431]. Recently Chevalier-Mariette et al. reduced the CpG content of the lacZ gene, resulting in an improved transgene that was significantly less sensitive to methylation induced silencing [429]. Interestingly, however, LacZ transgene arrays are able to induce transgene silencing in *Drosophila*, which only employ low levels of DNA methylation, indicating that alternative mechanisms of silencing may exist [432-434].

Second, Garrick et al. utilised transgenic lines with copy numbers in excess of 100. The highest copy number line reported herein contained only 20 transgene copies. This relatively small repeated array may be too small to induce repeat-induced gene silencing. Analysis of 1.3loxP transgenic mice containing larger transgene copy numbers may reveal a different phenotype.

Third, Garrick et al. did not address the potential presence of transgene rearrangements in the high copy lines. However, one of the low copy number lines produced after Cre mediated recombination contained a single rearranged copy of the α globin-lacZ transgene that did not express to a detectable level. This rearrangement must have been present in the high copy parental line, potentially accounting for some of the variegation noted in this line. In contrast the 'reduced copy number' line that displayed a less variegated phenotype did not contain rearranged transgene copies.

Finally, it is important to note that not all high-copy hCD2 transgenic lines display variegation even when they posses a disabled/truncated LCR (D. Kioussis unpublished observation). Thus, it is unlikely that variegation of hCD2 is a direct consequence of multiple transgene copies.

3.6.2 Imprinting in the 1.3lox lines

Flow cytometry revealed that in a number of 1.3loxP lines transgene expression was sensitive to the parental origin of the transgene. When the transgene was of a maternal origin the level of expression was lower and the level of variegation higher than when the transgene was of a paternal origin. Interestingly, transgene expression in each of these lines was not dependent upon the sex of the mouse being analysed. Furthermore, the imprinting event is reversed or erased on subsequent passage through the male germline. This characteristic is normally associated with 'classical imprinting', suggesting that similar mechanisms may be involved in imprinting of the 1.3loxP transgenes.

Imprinting of hCD2 transgenes has never been noted before and the mechanisms responsible for directing this imprinting are not understood. However, variegating lines that include intron 4 have not been analysed before. It therefore seems possible that this imprinting is influenced by DNA sequences residing within intron 4. Interestingly, all of the imprinted lines contain transgene copies integrated in the head to head orientation. However, it is not clear how this could direct imprinting.

The mechanisms responsible for directing this imprinting have yet to be investigated. However, a number of potential mechanisms can be envisaged. One possible explanation is that the transgenes are all integrated onto the X chromosome. In female 1.3loxP transgenic mice X chromosome inactivation would result in the transgene silencing in approximately 50% of the cells. However, breeding analysis rules out the possibility of X chromosome integrations (i.e. male mice passed the transgene to both male and female offspring). Alternatively, the hCD2 transgenes in the 1.3lox mice could be integrated proximal to naturally imprinted genes. However, as the imprinted phenotype is noted in 4 independently generated lines this seems improbable. An alternative possibility is that the 1.3loxP transgene is sensitive to the levels of a factor involved in imprinting within the germline. DNA methylation is the most likely candidate for mediating this imprinting, for a number of reasons. First, DNA methylation has been associated with transcriptional inactivity [435]. Second, DNA methylation has been implicated in the regulation of naturally imprinted genes [425]. Third, DNA methylation is stably preserved through cell division by maintenance methyltransferases [436]. Fourth, the degree of DNA methylation is different in spermatogenesis and oogenesis [437]. Fifth, the genome is extensively demethylated and then specifically remethylated during gametogenesis, providing a mechanism by

which the imprinted state could be reset [437]. Analysis of 1.3loxP mice containing a paternally or maternally derived transgene may therefore reveal a differential methylation of the hCD2 transgene. As hCD2 has never shown imprinted expression before, it is unlikely that this finding represents a normal mode of hCD2 regulation. It is more likely that this imprinting reflects a general differential epigenetic modification of the germline DNA between developing oocytes and spermatocytes. It remains unclear for the moment why the 1.3loxP transgene should be so susceptible to this differential epigenetic modification. This result is particularly important as it highlights the need to always consider the parental origin of the transgene, even when comparing mice in which the transgene is in the same site of integration.

Imprinted transgene variegation has previously been described [438-441]. For example, Kearns et al. generated a number of transgenic mouse lines carrying a ζ-globin-lacZ transgene (containing the human ζ-globin promoter, the lacZ reporter gene the erythroid-specific enhancer HS-40) [440]. In one line, passage of the transgene through the female germline resulted in silenced or reduced transgene expression (both the level of expression per cell and the percentage of expressing cells were reduced) [440]. In addition, silencing could be reversed by subsequent passage of the transgene through the male germline, which is consistent with classical imprinting [440]. The mechanisms responsible for this imprinting were not determined. However, female transmission was associated with DNA hypermethylation of the transgene. Interestingly, analysis of the DNA sequences proximal to the site of integration failed to reveal any allele-specific differential methylation in wild-type cells [440]. This suggests that the imprinting did not result from transgene integration into an imprinted locus.

Chapter Four

Transgenic Lines Expressing iCre

The work described in this chapter was done in collaboration with

Jasper de Boer

4.1 Generation of iCre transgenic mice

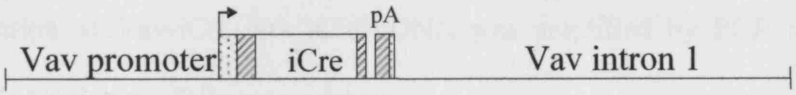
To enable the deletion of HSS3 at different developmental time points transgenic mice expressing Cre recombinase at different stages of haematopoietic development were required. Thus, in order to generate a system for conditional gene targeting within the myeloid and lymphoid lineages two different regulatory elements where chosen to direct expression of the codon-improved Cre recombinase (iCre) in transgenic mice. The iCre gene was chosen due to its adaptation for use in mammalian systems [442]. To generate iCre, the bacteriophage Cre gene was modified by silent base mutations so that it conforms to the mammalian codon usage and to eliminate cryptic splice sites [442]. The gene sequence was further modified to reduce the high CpG content of the original prokaryotic sequence, thereby reducing the chances of epigenetic silencing in mammals [442]. Finally an optimal Kozak consensus sequence was included as well as additional valine codon in front of the simian virus 40 T nuclear localisation signal [442].

Vav1 is expressed in all haematopoietic cells and plays an important role in both T cell development and T cell receptor mediated signalling (for a recent review see [443]. The vav expression vector (HS21/45 vav-hCD4) has previously been shown to drive expression of a hCD4 transgene throughout the entire haematopoietic compartment [444]. Thus, to restrict conditional targeting to the haematopoietic system, the hCD4 gene was removed from the vav expression vector and replaced with the iCre cDNA (figure 10). To further restrict expression to lymphocytes, the iCre cDNA was inserted into the hCD2 VA vector, which has previously been shown to drive high levels of transgene expression in T cells and occasionally in B cells, but not in other tissues (figure 10) [445]. As a control, the GFP open reading frame was inserted directly into the hCD2 VA cassette (figure 10).

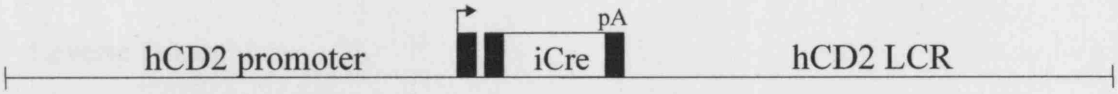
Figure 10. Cre expression constructs.

The iCre cDNA [442] was inserted into the HS21/45 vav-hCD4 expression cassette (Vav-iCre) [444]. The untranslated portion of the Vav1 exon 1 is represented by a light grey box. The SV40 splice sites and poly-adenylation signal (pA) are represented by dark grey boxes. The iCre and GFP cDNAs were inserted into the hCD2 VA cassette [445] (hCD2-iCre and hCD2-GFP, respectively). Exon 1, exon 2 and the hCD2 poly-adenylation signal (pA) are represented by black boxes.

Vav-iCre

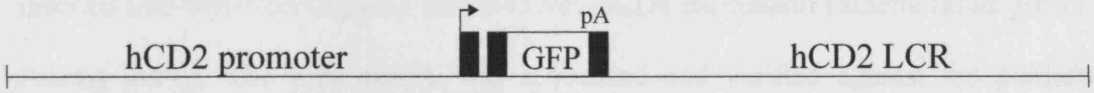


hCD2-iCre

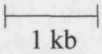


Construct generated by Alexandre Potocnik

hCD2-GFP



Construct generated by George Skavdis



4.2 Cloning

For the generation of hCD2-iCre, the iCre [442] coding sequence was subcloned into PCR-Script II, thereby altering the 5' and 3' multiple cloning sites. PCR-ScriptII-iCre was digested with NotI, filled in with Klenow polymerase and then digested with EcoRI to liberate the iCre cDNA. The iCre cDNA was then cloned into SmaI-EcoRI digested hCD2 VA vector.

For generation of Vav-iCre, the iCre cDNA was amplified by PCR of hCD2-iCre plasmid DNA using the following primers:

Forward Primer:

5'GCAGGCCCGTACGGCCATGGTGCCCAAGAAGAAGAGG3'

Reverse Primer:

5'GCATCAGCGCGGCCGCGTTTCAGTCCCCATCCTC3'

The PCR fragment containing the iCre cDNA was digested with SfiI and NotI and inserted into SfiI-NotI digested HS21/45 vav-hCD4 expression cassette (kind gift of J. Adams [444]). The iCre cDNA was sequenced and verified against the published sequence.

For generation of hCD2-GFP, GFP-S65T plasmid DNA was obtained from Quantum. The GFP cDNA was isolated by digestion with SacI-EcoRV before treatment with Klenow polymerase. The blunted GFP cDNA was then cloned into the SmaI site of the hCD2 VA vector.

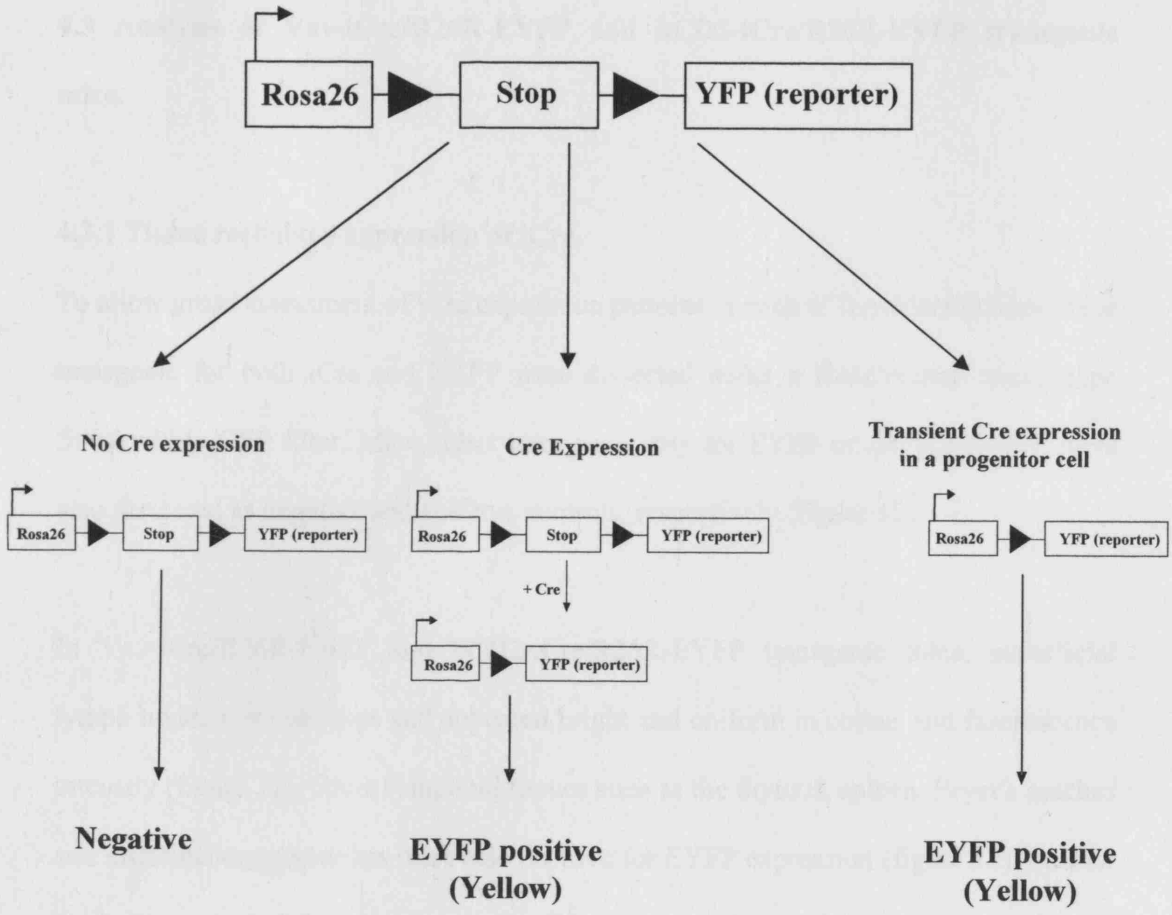
Each injection fragment was digested away from the vector sequences (KpnI-XbaI for hCD2-iCre, HindIII for Vav-iCre and KpnI-XbaI for hCD2-GFP) and purified. Pronuclei of fertilised oocytes from (CBA/Ca X C57F7BL/10) Fl mice were injected with the purified construct. The presence of the transgene in founders was determined by Southern blot analysis. After germline transmission a minimum of four independent mouse lines were established for each construct. Based on transgene expression profiles and the absence of disruption in the development of any haematopoietic lineage (see below), one of each of the Vav-iCre, hCD2-iCre and hCD2-GFP lines, were chosen for further analysis.

To determine the efficiency of iCre expression, Vav-iCre and hCD2-iCre transgenic mice were tested utilising two different systems. The first system relies upon a Cre recombinase-inducible EYFP reporter gene to visualise Cre expression in each transgenic line. This system allows detailed mapping of the cell types in which the recombinase is expressed. The second system applies negative selection pressure to Cre expressing cells through the use of a Cre-recombinase inducible diphtheria toxin (chapter 5). This system is able to reveal the presence of any inefficiency in either of the Cre expressing lines by providing a selective advantage to any non-expressing cells.

To map the Cre expression profile each transgenic line, Vav-iCre and hCD2-iCre transgenic mice were crossed to the homozygous R26R-EYFP reporter line [446]. The R26R-EYFP reporter line contains an EYFP transgene inserted into the ROSA26 locus by homologous recombination (figure 11). EYFP is only expressed when a floxed triple polyadenylation sequence, that prevents transcription from the ROSA26 locus, is excised by Cre mediated recombination [446]. Resulting EYFP expression can be

Figure 11. Fate mapping cells that have undergone recombination.

The R26R-EYFP reporter line has the EYFP cDNA inserted into the ubiquitously expressed ROSA26 locus. Transcription of the EYFP gene is prevented by a transcriptional stop sequence which is flanked by loxP sites (denoted by black triangles). Cre mediated recombination between the loxP sites results in deletion of the transcriptional stop sequence and activation of the EYFP gene. As the recombination event is a permanent modification all progeny will express EYFP, even if expression of Cre recombinase has ceased.


Rosa26
Stop
YFP (reporter)
No Cre expression
Cre Expression
Transient Cre expression in a progenitor cell
Rosa26
Stop
YFP (reporter)
+ Cre
Rosa26
YFP (reporter)
Rosa26
YFP (reporter)
Negative
EYFP positive (Yellow)
EYFP positive (Yellow)

detected by either flow cytometry or by fluorescence microscopy. Importantly, activation of the R26R-EYFP reporter gene is an irreversible event resulting in not only the permanent marking of cells that have undergone recombination but also in the marking of all subsequent progeny.

4.3 Analysis of Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice.

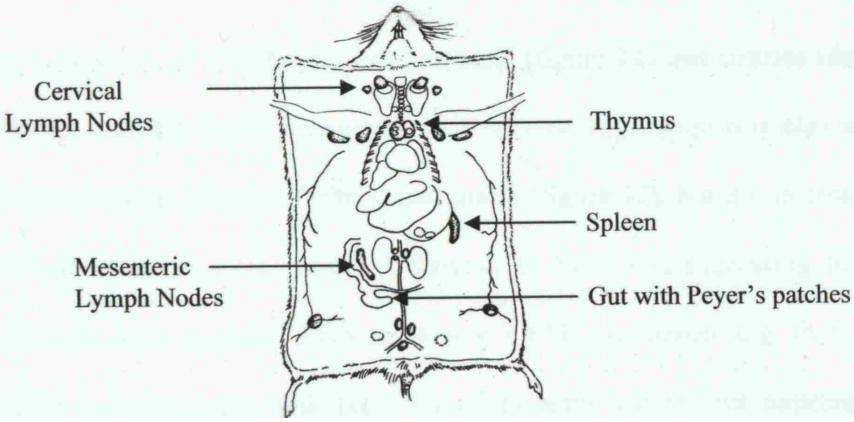
4.3.1 Tissue restricted expression of iCre.

To allow gross assessment of iCre expression patterns in each of the selected lines, mice transgenic for both iCre and EYFP were dissected under a fluorescence microscope fitted with a GFP filter. Mice either transgenic only for EYFP or for hCD2-GFP were also dissected as negative and positive controls, respectively (figure 12).

In Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice, superficial lymph nodes were obvious and appeared bright and uniform in colour and fluorescence intensity (figure 12). Other lymphoid tissues such as the thymus, spleen, Peyer's patches and intestinal cryptopatches were also positive for EYFP expression (figure 12). Further dissection revealed that the mesenteric (figure 12) and inguinal lymph nodes (data not shown) were positive for EYFP expression. Analysis of hCD2-GFP transgenic mice showed an almost identical distribution of fluorescence, as observed in Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice. In contrast, however, to the uniform expression as seen in Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice, the Peyer's patches from hCD2-GFP mice exhibited areas of high and low fluorescence intensity.

Figure 12. Analysis of Cre expression patterns in tissues of transgenic mice by fluorescence microscopy.

Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice and hCD2-GFP mice were dissected under a fluorescence microscope fitted with a GFP filter. An R26R-EYFP single transgenic mouse is also shown as a negative control. Similar results were obtained in at least five independent experiments.


Cervical
Lymph Nodes
Thymus
Spleen
Mesenteric
Lymph Nodes
Gut with Peyer's patches

Surprisingly, expression of EYFP was noted in testis (figure 12) and ovaries (data not shown) of Vav-iCre/R26R-EYFP transgenic mice. Mosaic expression was also seen in the testis of hCD2-iCre/R26R-EYFP transgenic males (figure 12), but not in ovaries of hCD2-iCre/R26R-EYFP females (data not shown). In both iCre-expressing lines all other tested tissues were negative for macroscopic EYFP expression (e.g. liver, lung, muscle and gut), indicating that, with exception of the testis and ovaries, expression of iCre was restricted to tissues of the haematopoietic and/or immune system. However, sporadic ectopic expression of EYFP was detected occasionally in Vav-iCre/R26R-EYFP transgenic mice (see below). Expression of GFP was not detected in testis or ovaries from hCD2-GFP mice (figure 12). No EYFP expression was detectable in any tissue analysed from R26R-EYFP single transgenic controls.

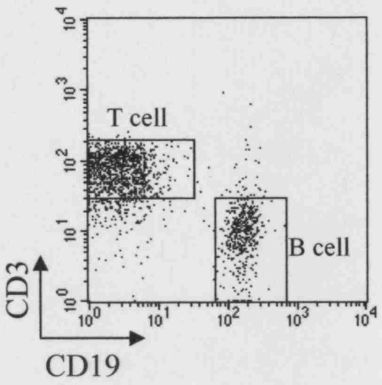
4.3.2 iCre expression in T cells

In order to identify the types of cells that have undergone iCre mediated recombination within the positive organs described above, cells isolated from lymph nodes were stained with anti-CD3 and anti-CD19 and analysed by flow cytometry (figure 13). Analysis of both Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice revealed that all peripheral $CD3^+$ T cells expressed EYFP. In addition, high levels of GFP expression were detected in all peripheral T cells from hCD2-GFP mice. Thymocytes were stained with anti-CD4 and anti-CD8 to distinguish the $CD4^+/CD8^-$ or $CD4^-/CD8^+$ single positive (SP) thymocytes as well as $CD4^+/CD8^+$ double positive (DP) thymocytes (figure 14). Detailed analysis revealed that Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice exhibited high levels of EYFP expression on approximately 100% of both SP and DP thymocytes. This finding indicates that iCre

Figure 13. Flow cytometric analysis of EYFP and GFP expression in lymph node derived lymphocytes.

Single cell suspensions from lymph nodes were prepared from Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice. These cells were stained with CD3 and CD19 antibodies to allow differentiation of the T cell ($CD3^+/CD19^-$) and B cell populations ($CD3^-/CD19^+$). Levels of EYFP or GFP expression on T cells and B cells are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the level of auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

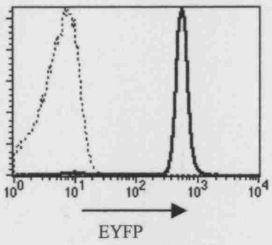
Lymph node

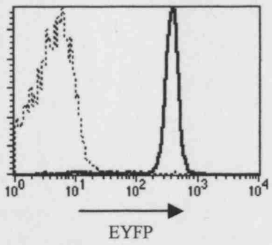


Peripheral T cells

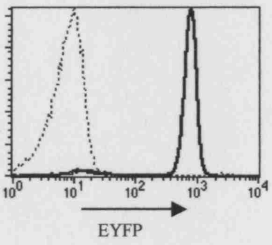
Peripheral B cells

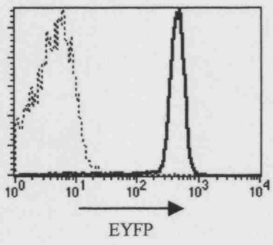
VAV-iCre



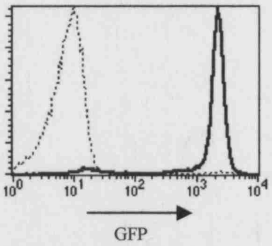


hCD2-iCre





hCD2-GFP



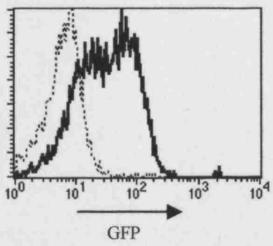
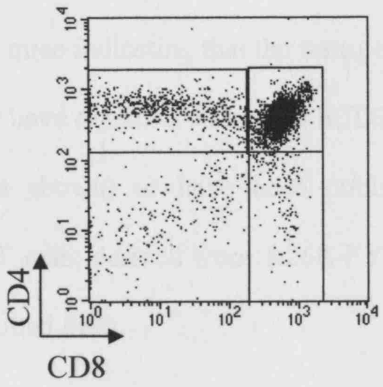
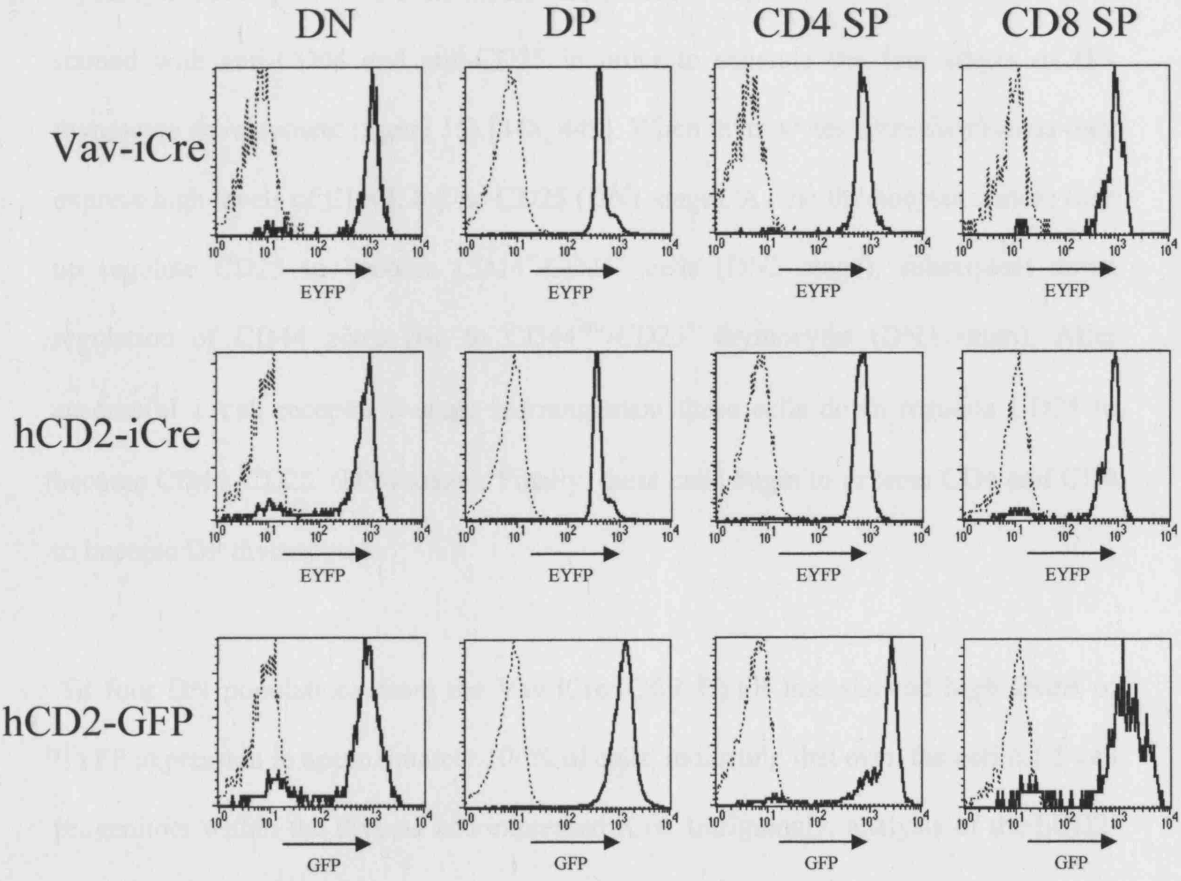


Figure 14. Flow cytometric analysis of EYFP and GFP expression in thymocytes.

Thymocytes from Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with CD4 and CD8 antibodies. Levels of EYFP or GFP expression on SP ($CD4^+/CD8^-$ or $CD4^-/CD8^+$), DP ($CD4^+/CD8^+$) and DN populations ($CD4^-/CD8^-$) are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the level of auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Thymus





was expressed in these cells or had been expressed at an earlier developmental time point. Identical analysis revealed a similar percentage of GFP expressing SP and DP thymocytes in hCD2-GFP mice indicating that the transgene is not subject to PEV. This is in contrast to results we have obtained with other hCD2-GFP transgenic lines that we have generated (data not shown) or have been published [447]. Importantly, no expression was seen on T cells derived from R26R-EYFP single transgenic controls (figure 13 and figure 14 dotted line).

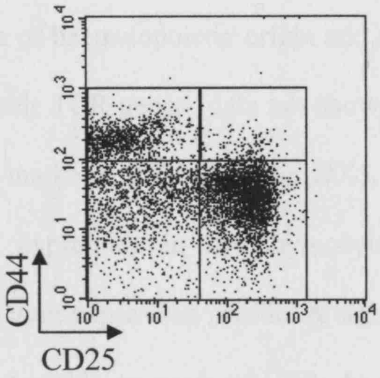
To determine precisely the time point at which iCre expression is initiated during thymocyte development in each of the lines, double negative (DN) thymocytes were stained with anti-CD44 and anti-CD25 in order to separate the four stages of DN thymocyte development (figure 15) [448, 449]. When thymocytes enter the thymus they express high levels of CD44, but no CD25 (DN1 stage). As the thymocytes mature they up regulate CD25 to become $CD44^{+}/CD25^{+}$ cells (DN2 stage); subsequent down regulation of CD44 gives rise to $CD44^{low}/CD25^{+}$ thymocytes (DN3 stage). After successful T cell receptor β-chain rearrangement these cells down regulate CD25 to become $CD44^{-}/CD25^{-}$ (DN4 stage). Finally, these cells begin to express CD4 and CD8 to become DP thymocytes.

All four DN populations from the Vav-iCre/R26R-EYFP line showed high levels of EYFP expression in approximately 100% of cells, indicating that even the earliest T cell progenitors within the thymus had expressed iCre. Intriguingly, analysis of the hCD2-iCre/R26R-EYFP line revealed a slightly different pattern of EYFP expression. A population of EYFP negative cells were detected in DN1, DN2, and DN3 stages, which

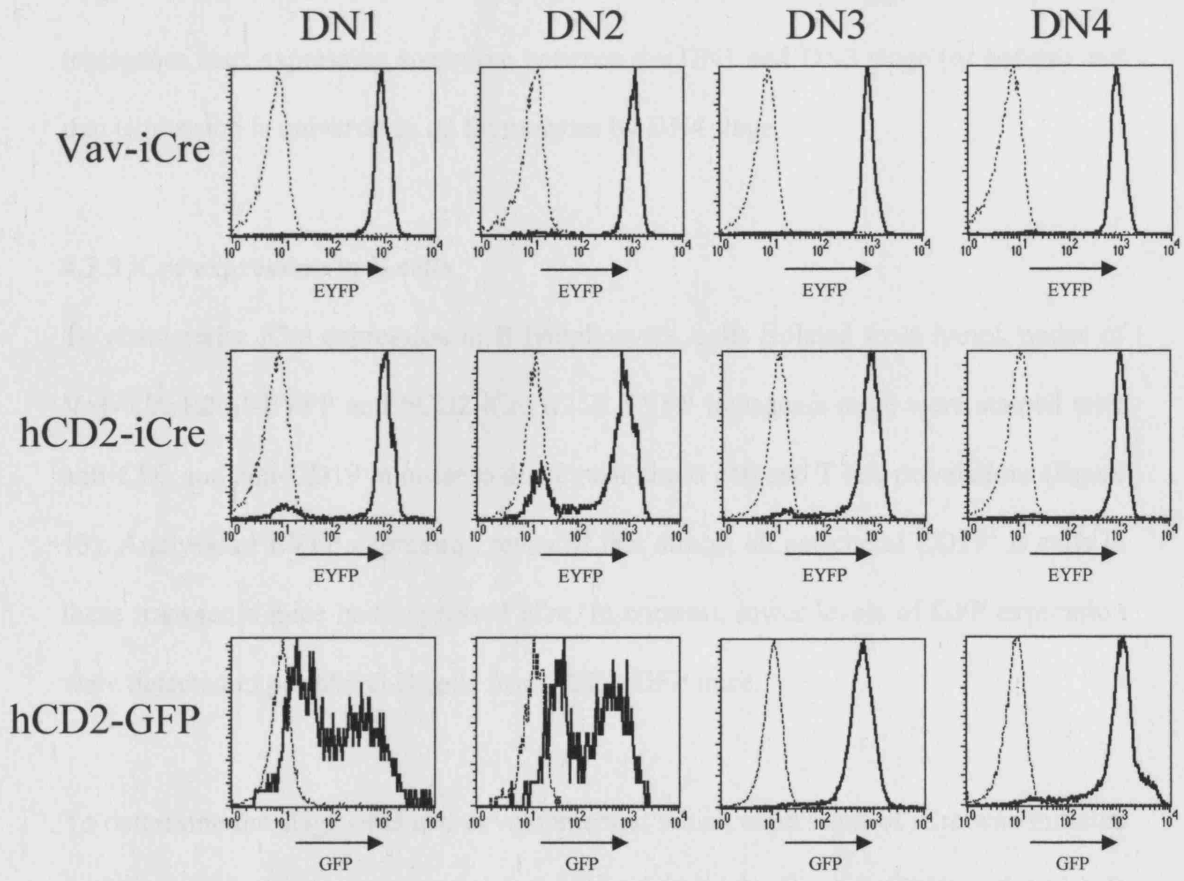
Figure 15. Flow cytometric analysis of EYFP and GFP expression during DN thymocyte development.

Single cell suspensions from thymocytes from Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were depleted of CD4 and CD8 expressing cells by complement treatment (see materials and methods). Remaining DP and SP thymocytes were excluded from analysis by staining with CD4, and CD8 antibodies. Resulting DN thymocytes were stained with anti-CD25 and anti-CD44 antibodies. Levels of EYFP or GFP expression at each DN stage are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the level of auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Double negative thymocytes



DN1	DN2
DN4	DN3



disappeared in the final DN4 stage. Phenotypic analysis of these EYFP negative cells revealed that they are positive for Thy-1 and CD45 but negative for CD3, indicating that these cells are of haematopoietic origin and are possibly DN thymocytes that have yet to rearrange their TCR genes (data not shown). The same cells were also negative for the following markers: B220, CD11c, DX5, Mac1 and Ter119 (data not shown). Analysis of GFP expression in DN thymocytes from hCD2-GFP transgenic mice revealed yet a different expression profile. A significant proportion of DN1 thymocytes did not express detectable amounts of GFP. As the thymocytes moved through to the DN2 stage the proportion of GFP expressing cells increased until the DN3 and DN4 stages, where 100% of the cells expressed GFP. These data suggest that the hCD2 transgenes start expressing sometime between the DN1 and DN3 stage (or before) and that expression is universal in all thymocytes by DN4 stage.

4.3.3 iCre expression in B cells

To characterise iCre expression in B lymphocytes, cells isolated from lymph nodes of Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice were stained with anti-CD3 and anti-CD19 in order to distinguish the B cell and T cell populations (figure 13). Analysis of EYFP expression revealed that almost all peripheral $CD19^+$ B cells in these transgenic mice had expressed iCre. In contrast, lower levels of GFP expression were detected in peripheral B cells from hCD2-GFP mice.

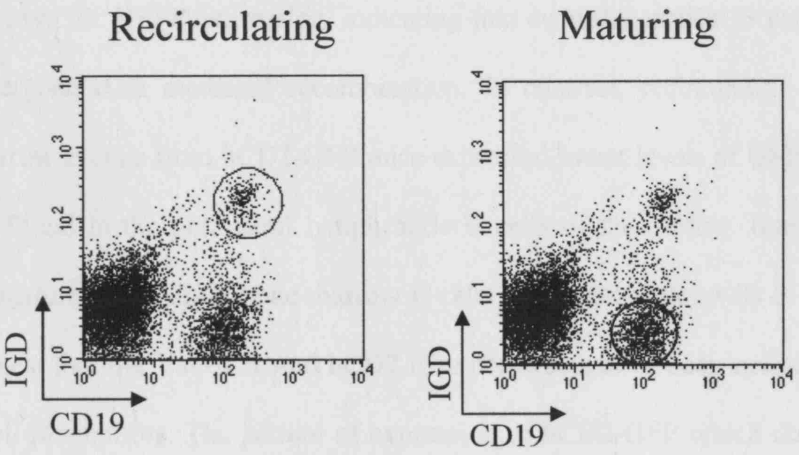
To determine the stage of B cell development at which expression of iCre was initiated in this lineage, bone marrow was isolated and stained with anti-CD19 and anti-IgD antibodies (figure 16). $CD19^+/IgD^+$ bone marrow cells are mature recirculating B cells, whereas $CD19^+/IgD^-$ bone marrow cells represent stages of B cell development prior to

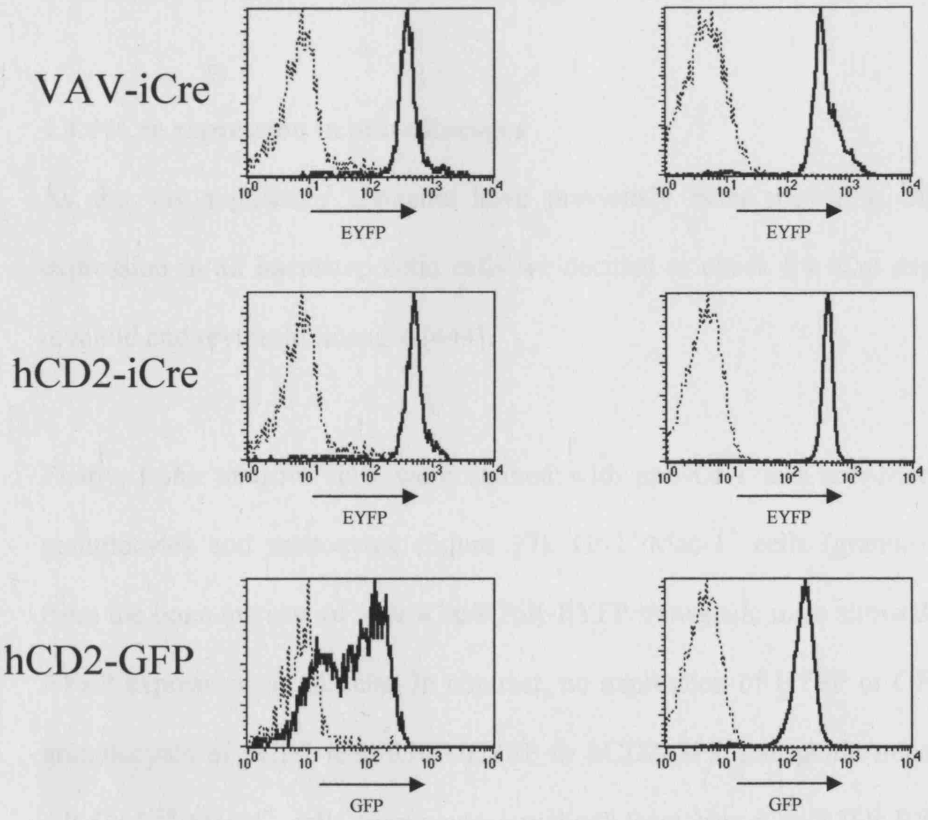
Figure 16. Flow cytometric analysis of EYFP and GFP expression in bone marrow derived B cells.

Single cell suspensions from bone marrow from Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with CD19 and IgD antibodies. Levels of EYFP or GFP expression on mature recirculating $CD19^{+}/IgD^{+}$ B cells and $CD19^{+}/IgD^{-}$ immature B cells are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the level of auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Bone marrow

B cells





migration into the periphery. All mature $CD19^+/IgD^+$ and $CD19^+/IgD^-$ bone marrow derived B cells in Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice were positive for EYFP expression, indicating that even the earliest B cell progenitors had undergone iCre mediated recombination. In contrast, recirculating $CD19^+/IgD^+$ bone marrow B cells from hCD2-GFP mice expressed lower levels of GFP, comparable to those found in the peripheral lymph node B cells of these mice. Interestingly, the more immature $CD19^+/IgD^-$ bone marrow B cells expressed high levels of GFP. These data suggest that the Vav-iCre and hCD2-iCre constructs have been active in all B cell and T cell progenitors. The pattern of expression of hCD2-GFP which does not detect the history of expression, but current expression, indicates that the transgenic construct is active in early stages of B cell development but its expression is diminished as the B cells mature leading to a picture resembling PEV in this cell lineage.

4.3.4 iCre expression in other lineages

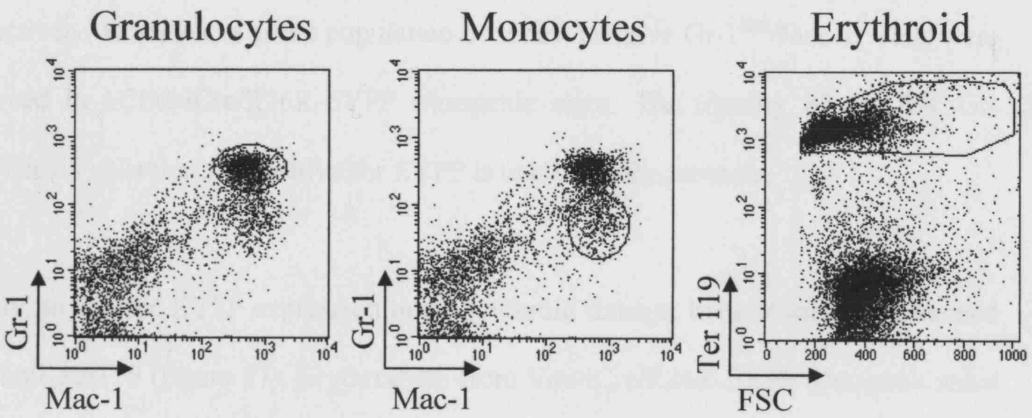
As the vav regulatory elements have previously been shown to direct transgene expression in all haematopoietic cells we decided to check for iCre expression in the myeloid and erythroid lineages [444].

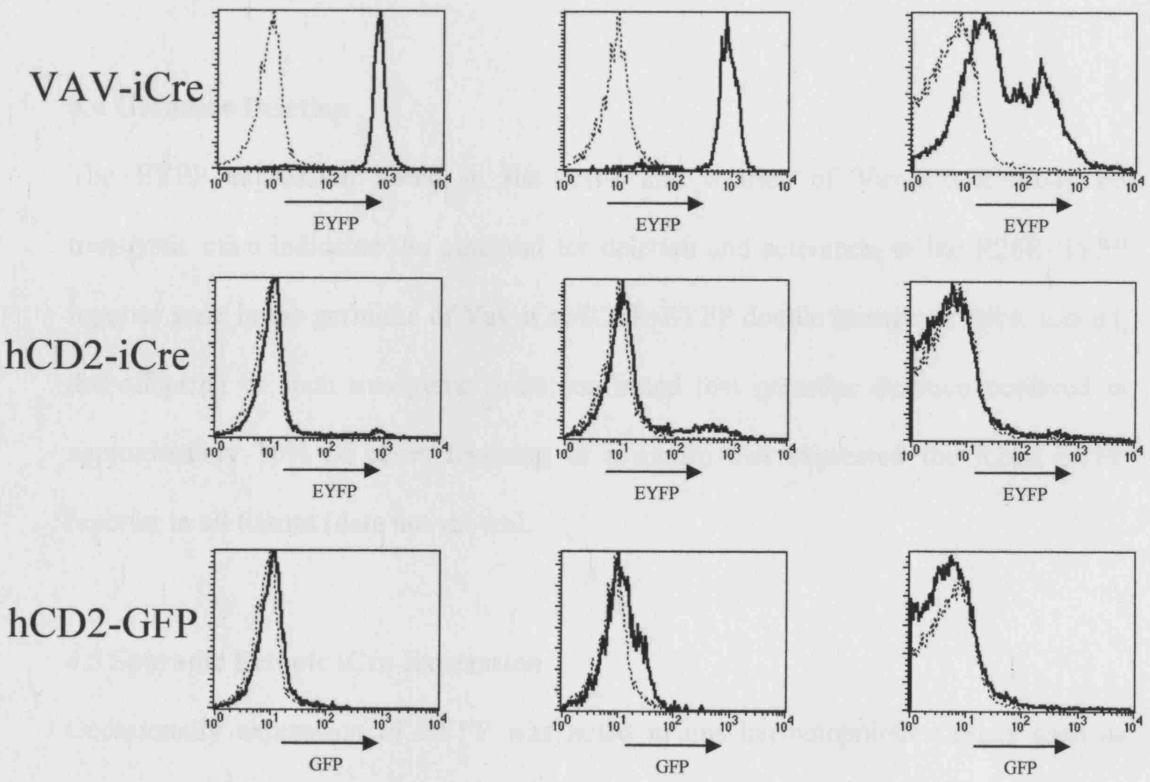
Firstly, bone marrow cells were stained with anti-Gr-1 and anti-Mac-1 to identify granulocytes and monocytes (figure 17). $Gr\text{-}1^+/Mac\text{-}1^+$ cells (granulocyte enriched) from the bone marrow of Vav-iCre/R26R-EYFP transgenic mice showed high levels of EYFP expression in all cells. In contrast, no expression of EYFP or GFP was seen in granulocytes of hCD2-iCre/R26R-EYFP or hCD2-GFP transgenic mice, respectively. All $Gr\text{-}1^{low}/Mac\text{-}1^+$ cells (monocyte enriched) from Vav-iCre/R26R-EYFP transgenic mice showed high levels of EYFP. No substantial EYFP or GFP expression was seen in

Figure 17. Flow cytometric analysis of EYFP and GFP expression in bone marrow derived myeloid and erythroid cells.

Single cell suspensions from bone marrow from Vav-iCre/R26R-EYFP double transgenic mice, hCD2-iCre/R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with Gr-1 and Mac-1 or Ter119 antibodies. Levels of EYFP or GFP expression on granulocytes Gr-1^{+}/Mac-1^{+}, monocytes Gr-1^{+}/Mac-1^{low} or erythrocytes are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the level of auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Bone marrow


Granulocytes
Monocytes
Erythroid
Gr-1
Mac-1
Gr-1
Mac-1
Ter 119
FSC


VAV-iCre
hCD2-iCre
hCD2-GFP
EYFP
EYFP
EYFP
EYFP
EYFP
EYFP
GFP
GFP
GFP

monocytes from either hCD2-iCre/R26R-EYFP or hCD2-GFP transgenic mice, respectively. However, a small population of EYFP positive Gr-1^{low}/Mac-1^{+} cells were observed in hCD2-iCre/R26R-EYFP transgenic mice. The identity of the few Gr-1^{low}/Mac-1^{+} cells that are positive for EYFP is unclear at the moment.

Finally, to analyse EYFP expression in the erythroid lineage, bone marrow was stained with anti-Ter119 (figure 17). Erythrocytes from Vav-iCre/R26R-EYFP transgenic mice were seen to express EYFP, although the levels of expression varied significantly between cells. No EYFP or GFP expression was seen in erythrocytes from hCD2-iCre/R26R-EYFP or hCD2-GFP transgenic mice, respectively.

4.4 Germline Deletion

The EYFP expression noted in the testis and ovaries of Vav-iCre/R26R-EYFP transgenic mice indicated the potential for deletion and activation of the R26R-EYFP reporter gene in the germline of Vav-iCre/R26R-EYFP double transgenic mice. Indeed, the offspring of such transgenic mice confirmed that germline deletion occurred in approximately 70% of cases resulting in a mouse that expressed the R26R-EYFP reporter in all tissues (data not shown).

4.5 Sporadic Ectopic iCre Expression

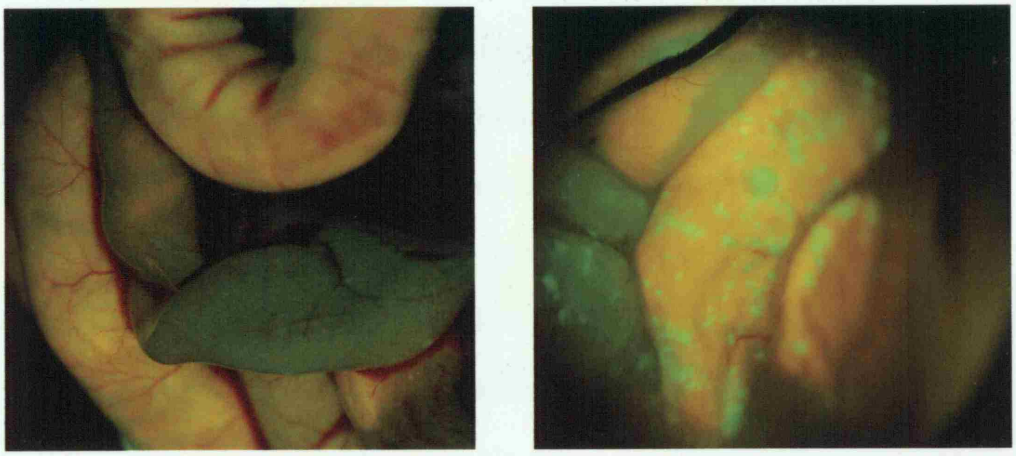
Occasionally expression of EYFP was noted in non-haematopoietic tissues such as liver, heart and gut of Vav-iCre/R26R-EYFP transgenic mice, indicating that the Vav-iCre transgene has been expressed in these cells (figure 18A). One possible explanation is that $EYFP^{+}$ T cells have infiltrated into these tissues resulting in the apparent ectopic expression. To test this hypothesis cells were isolated from the intestines of mice

Figure 18. Sporadic ectopic expression in Vav-iCre/R26R-EYFP double transgenic mice.

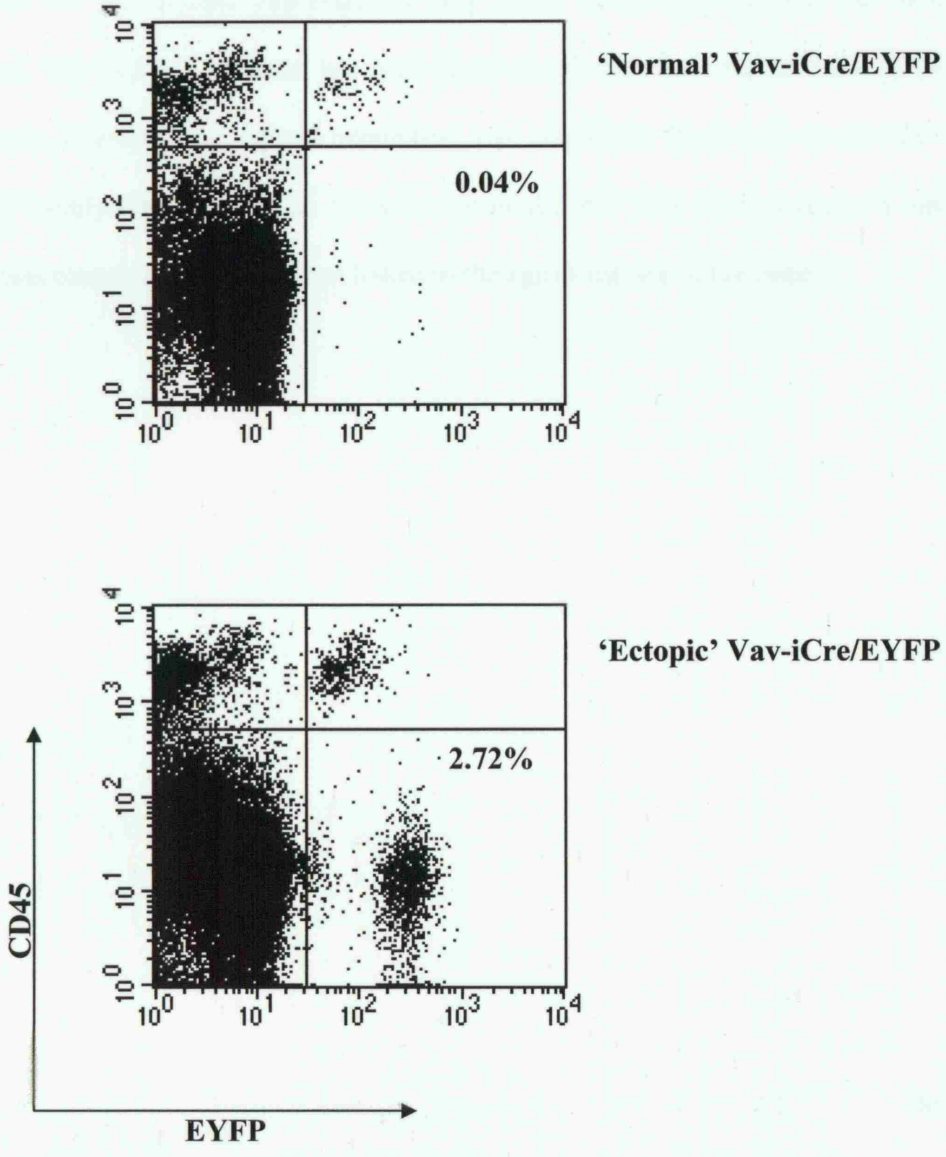
(A) Vav-iCre/R26R-EYFP double transgenic mice displaying the normal pattern of expression (left panel) and Vav-iCre/R26R-EYFP double transgenic mice displaying ectopic expression (right panel) were dissected under a fluorescent microscope fitted with a GFP filter.

(B) Single cell suspensions from the small intestine of Vav-iCre/R26R-EYFP double transgenic mice displaying the normal pattern of expression and Vav-iCre/R26R-EYFP double transgenic mice displaying ectopic expression were stained anti-CD45, which is a pan-haematopoietic marker. The percentage of $CD45^-$ cells expressing EYFP are indicated.

A.


'Normal'
'Ectopic'

B.


'Normal' Vav-iCre/EYFP
0.04%
'Ectopic' Vav-iCre/EYFP
2.72%
CD45
EYFP

displaying either normal or ectopic expression of EYFP (see materials and methods). The isolated cells were stained with anti-CD45, which is a pan-haematopoietic surface maker, and analysed for expression of EYFP (figure 18B). In Vav-iCre/R26R-EYFP mice with expected patterns of expression less than 0.1% of $CD45^-$ cells (cells of non-haematopoietic origin) express EYFP. In contrast, in Vav-iCre/R26R-EYFP mice displaying ectopic expression, EYFP was detected on approximately 3% of cells that were $CD45^-$, and therefore non-haematopoietic. This indicates that the reporter gene expression in the intestines of these mice results from a true ectopic expression of EYFP and not from an infiltration of $EYFP^+$ T cells.

The reason that this ectopic expression is only noted infrequently has not yet been determined. The hCD2-iCre strain has been backcrossed onto both the B10 and CBA backgrounds. Interestingly, ectopic expression was noted less frequently on the B10 (table 1A) background than on the CBA background (table 1B). Further analysis has revealed that ectopic expression is not linked to the age or the sex of the mice.

Table 1. Numbers of Vav-iCre/R26R-EYFP double transgenic mice displaying ectopic EYFP expression.

(A) Table showing the numbers of male and female Vav-iCre/R26R-EYFP double transgenic mice of different ages displaying ectopic versus expected EYFP expression on the B10 background.

(B) Table showing the numbers of male and female Vav-iCre/R26R-EYFP double transgenic mice of different ages displaying ectopic versus expected EYFP expression on the CBA background.

A.

Numbers of Ectopic Vs Normal Mice on the B10 Background					
Age (Days)	Ectopic		Normal		Ectopic/Total
	M	F	M	F	
32	1	0	1	2	1/4
39	0	1	2	3	1/6
46	0	0	0	3	0/3
47	0	0	4	1	0/5
54	1	0	1	2	1/4
62	0	0	1	0	0/1
74	0	1	1	0	1/2
94	1	0	5	0	1/6
100	0	1	3	1	1/5
109	1	0	3	2	1/6
118	1	0	2	0	1/2
Combined Total	**5**	**3**	**23**	**14**	**8/45**

B.

Numbers of Ectopic Vs Normal Mice on the CBA Background					
Age (Days)	Ectopic		Normal		Ectopic/Total
	M	F	M	F	
28	0	1	0	6	1/7
33	2	0	0	0	2/2
34	2	0	1	0	2/3
49	0	1	1	2	1/4
59	0	2	1	0	2/3
64	0	1	0	1	1/2
80	1	0	0	2	1/3
95	1	1	1	0	2/3
109	1	0	0	0	1/1
121	0	0	0	2	0/2
Combined Total	**7**	**6**	**4**	**13**	**13/30**

4.6 Discussion

To allow the deletion of HSS3 at different developmental time points two independent Cre expressing transgenic lines were generated [450]. In the first line, expression of iCre is driven by the vav regulatory elements, which direct expression in all haematopoietic lineages, presumably including the haematopoietic stem cells. In the second line, expression of iCre is driven by the hCD2 promoter and LCR which direct expression in the lymphoid compartment.

4.6.1 Tissue Restricted Expression of Cre

Fluorescence microscopy revealed that EYFP was expressed in all lymphoid tissues of both Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP double transgenic mice, indicating that Cre was expressed in these tissues to a level sufficient to mediate the deletion of the floxed R26R-EYFP stop sequence. Interestingly, EYFP expression was detected in the testis and ovaries of Vav-iCre/R26R-EYFP transgenic mice. Although the endogenous vav gene is expressed in the testis, the regulatory elements responsible for directing this expression are not included in the expression cassette used to generate the Vav-iCre transgenic mice [451]. In addition, a proportion of cells within the testis of hCD2-iCre/R26R-EYFP transgenic mice expressed EYFP. In contrast, expression of hCD2 has never been noted in the testis of hCD2 transgenic mice, nor has expression of GFP been noted in the testis of hCD2-GFP mice.

4.6.2 Germline Deletion

The EYFP expression noted in the testis and ovaries of Vav-iCre/R26R-EYFP transgenic mice indicated the potential for deletion and activation of the R26R-EYFP

reporter gene in the germline of Vav-iCre/R26R-EYFP double transgenic mice. Indeed, the offspring of such transgenic mice confirmed that germline deletion occurred in approximately 70% of cases. It is unlikely that the testis expression is due to the site of transgene integration, as it has been observed in the two independent lines presented here, as well as other Vav-iCre and hCD2-iCre lines that we have generated. It seems more likely that iCre contains DNA sequences that are able to activate transcription of the Vav-iCre and hCD2-iCre transgenes in the testis of these mice. This is not an unreasonable suggestion as for a long time prokaryotic sequences have been known to cause dysregulation in transgenic animals [452]. For example, the Thy-1 regulatory elements have previously been reported to drive tissue-specific, position-independent transgene expression [453]. However, when the same regulatory elements are linked to a lacZ reporter gene high levels of transgene expression are observed in the testis [453]. The authors suggest that the lacZ coding sequence may either disrupt the normal Thy-1 regulatory signals or create new regulatory elements that are able to drive expression outside of the normal cell types [453].

Perhaps more relevant to the work described herein is the recent publication by Eckardt el. who detailed the levels of ectopic expression in 6 unrelated Cre expressing transgenic mouse lines [454]. The results obtained indicate that Cre recombinase has a high potential to mediate ectopic transgene expression.

4.6.3 Sporadic Ectopic Cre Expression

In addition to the consistent expression of EYFP noted in the testis and ovaries of Vav-iCre/R26R-EYFP transgenic mice ectopic EYFP expression was occasionally noted in non-haematopoietic tissues such as gut and heart of Vav-iCre/R26R-EYFP transgenic

mice, indicating that the Vav-iCre transgene has been expressed in cells of these organs. Flow cytometry indicated that this ectopic expression did not result from an infiltration of EYFP positive T cells into the respective tissue but rather a true ectopic expression of the Vav-iCre transgene in non-haematopoietic cell types. The reason that this ectopic expression is only noted infrequently has not yet been determined. Originally the founder mice were of [(CBA/ca x C57BL/10) x (CBA/ca x C57BL/10)] genetic background, but to establish the line they have been backcrossed onto the CBA and B10 backgrounds. Therefore, the sporadic nature of the ectopic expression may represent a segregation of genetic elements, which may influence (through an undefined mechanism) how tightly the transgene is regulated. Some support for this hypothesis comes from the finding that ectopic expression is noted less frequently on the B10 background than the CBA background. Perhaps relevant to this notion, genetic background has previously been reported to influence the extent of transgene variegation [455].

In addition to the sporadic ectopic expression noted in the Vav-iCre line presented above, a number of other Vav-iCre lines that we have generated have exhibited mosaic expression of Cre in several non-haematopoietic tissues, including gut, brain, and muscle. A vavCre line has been previously described and expression of the transgene was noted in all haematopoietic tissues [456]. Interestingly, the authors noted expression in the vascular endothelium of gut, brain, heart and testis. Together these results support the notion mentioned above that some prokaryotic transgenes are able to cause aberrant transcriptional regulation of transgenes and neighbouring genes at the site of integration.

In addition to driving the inappropriate activation of transgenes, prokaryotic sequences are frequently associated with induction of gene silencing. One potential explanation for this phenomenon is that prokaryotic sequences are frequently enriched in CpG dinucleotides when compared to eukaryotic sequences. In mammals such CpG repeats represent targets for DNA methyl transferases and DNA methylation is associated with gene silencing [429]. Indeed, the CpG rich lacZ gene has been shown to deregulate a number of linked genes [429, 452]. For example, as previously mentioned, Ogilvy et al. reported that the vav regulatory elements are able to drive expression of a linked hCD4 reporter gene in all cells of haematopoietic origin [444]. In contrast, when the vav regulatory elements were used to drive lacZ expression, all transgenic lines exhibited variegated expression of the reporter, and the percentage of β-gal$^+$ cells decrease as the animals aged [457].

Evidence for Cre-induced transcriptional deregulation is manifest in a number of hCD2-iCre transgenic lines that we have produced. The hCD2 LCR is able to direct tissue-specific expression of mammalian transgenes in a position-independent, copy-number dependent manner. However, in a number of hCD2-iCre lines, expression of Cre was either mosaic or not detectable. This suggests that in these mice the Cre gene was able to overcome the powerful regulatory elements of the hCD2 LCR and induce variegation, possibly by enforcing heterochromatinisation of the transgenic locus at the site of integration.

A transgenic line expressing Cre under the control of the hCD2 regulatory elements has been previously described, and according to expectation Cre activity was detected only in lymphoid tissues [458]. However, the detection system used in this study did not

allow the assessment of Cre expression in individual cells and therefore, the issue of variegation was not addressed.

In addition to the ability of Cre to induce transcriptional deregulation, we also noted toxic effects of Cre expression. The Vav-iCre and hCD2-iCre transgenic lines presented in this study showed no detectable differences in thymic cellularity or development when compared to non-transgenic littermate controls. However, in a number of other hCD2-iCre lines and one Vav-iCre line the size of the thymus was decreased approximately tenfold when compared to non-transgenic littermates. High levels of Cre expression have been reported to be toxic in eukaryotes, possibly due to chromosomal rearrangements caused by recombination between cryptic 'pseudo-loxP' sites naturally found within the genome [459]. Alternatively, integration of the transgene into the genome of these mice may have disrupted a gene important for thymic development. However, this possibility seems unlikely as the phenomenon was noted in a number of independent transgenic lines.

4.6.4 Pattern of Cre Expression in T Cells

Flow cytometry revealed that EYFP was expressed in almost all peripheral SP T cells and all SP and DP thymocytes from Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic mice. Previously it has been shown that GFP transfected haematopoietic cell lines undergoing apoptosis or necrosis display decreased GFP fluorescence. Furthermore, the decreased EGFP fluorescence appears to be an early and very sensitive marker of cell death [460]. Interestingly, in such cases the amount of fluorescent protein is similar in living and dying cells ruling out the possibility that fluorescent proteins can leak out of cells or are degraded [461]. It has therefore been

suggested that fluorescent proteins are quenched by modifications in the intra-cellular milieu. Thus, dying cells expressing fluorescent proteins lose their fluorescence and can cause false negative results when not excluded from the analysis. We therefore believe that the small population of EYFP negative cells possibly represent damaged cells that have lost EYFP fluorescence, rather than cells that have not expressed Cre. It should be noted that if cells were not kept under optimal conditions (i.e. on ice) or were subjected to treatment with red blood cell lysis buffer (NH_4Cl), a substantial number of T cells became EYFP negative. It is also interesting to note that T cells were more prone to loss of fluorescence when compared to B cells, possibly reflecting the rate of death for these two different cell types *in vitro*.

Flow cytometric analysis of DN thymocytes from Vav-iCre/R26R-EYFP transgenic mice revealed that even the earliest progenitors to enter the thymus express the R26R-EYFP reporter gene. In contrast, analysis of hCD2-iCre mice revealed a small population of EYFP negative cells within the DN1, DN2 and DN3 gates. In hCD2-GFP mice, the amount of GFP negative cells decreased from the DN1 stage until the DN4 stage where all cells were GFP positive. Fewer GFP negative cells were found in DN1 from the hCD2-iCre/R26R-EYFP mice compared to hCD2-GFP mice. These differences could be due to a number of factors. First the hCD2-iCre transgenic locus might be integrated in a more accessible chromosomal locus compared to the hCD2-GFP locus, allowing expression of the hCD2-iCre transgene to be initiated earlier. This hypothesis seems plausible as different hCD2 lines begin to express hCD2 at slightly different stages within DN development (D. Kioussis unpublished observations). These differences could also result from the mechanism used to report expression of each construct. Expression of Cre results in the deletion of the R26R-EYFP stop signal,

which is a permanent modification that results in the expression of EYFP. All subsequent daughter cells therefore express EYFP and at levels determined by the ROSA26 locus. In contrast, expression of GFP is directly under the control of the hCD2 regulatory elements and will only be expressed when the hCD2-GFP transgene is actively being transcribed and at levels determined by the hCD2 regulatory elements. It is therefore possible that iCre had been expressed transiently in a proportion of cells at an early developmental stage prior to entering the thymus at the DN1 stage. Such expression would not be detected in the hCD2-GFP line.

4.6.5 Patterns of iCre Expression in B Cells

In both Vav-iCre/R26R-EYFP and hCD2-iCre/R26R-EYFP transgenic lines, all peripheral B cells express EYFP indicating that Cre mediated deletion had occurred early in B cell development. Importantly, no disturbances in the B:T cell ratio were observed indicating that Cre expression had not perturbed development of either lineage. In contrast, lower levels of GFP were expressed in peripheral B cells of hCD2-GFP mice. The apparent difference in expression profiles between the hCD2-iCre and hCD2-GFP lines is likely to be dependent upon the mechanism used to report transgene expression. Cre expression results in the deletion of the R26R-EYFP stop signal, which is a permanent modification. All subsequent daughter cells inherit the deletion and express EYFP, which can therefore be viewed as a history of Cre expression. In contrast, expression of GFP is under direct control of the hCD2 regulatory elements and will only be expressed when the hCD2-GFP transgene is actively transcribed. Taken together, these results suggest that hCD2 may be expressed early in B cell development, but its expression is shut down on B cell maturation.

4.6.6 Patterns of iCre Expression in Other Haematopoietic Lineages

It has previously been reported that the vav regulatory elements are able to direct transgene expression in all haematopoietic cell types [444]. Accordingly, in Vav-iCre/R26R-EYFP transgenic mice expression of EYFP was observed in all granulocytes, monocytes and erythrocytes, indicating that Cre had been expressed in all these cells and possibly their progenitors. Expression of EYFP in erythrocytes varied significantly from cell-to-cell within the same sample. This is probably due to the progressive loss of EYFP protein after the enucleation of these cells. Alternatively, the haemoglobin present within these cells may be quenching the EYFP fluorescence. In agreement with the known pattern of hCD2 expression in transgenic mice, no expression of EYFP or GFP was detected in these cell types in either hCD2-iCre/R26R-EYFP or the hCD2-GFP transgenic mice, respectively.

In summary, these two Cre expressing transgenic lines will be valuable in producing gene knockouts restricted to the haematopoietic system. The Vav-iCre will allow reliable deletion of floxed targets (such as HSS3) throughout the entire haematopoietic compartment, whereas the hCD2-iCre line will allow targeting to be restricted to T cells and B cells alone. Importantly, as expression of Cre is minimal outside of these tissues, apart from the testis and ovaries, all other cell types will retain the floxed target, minimising secondary effects of the deletion. Finally, complete deletion within each tissue will remove the phenotypic complications caused by the presence of non-deleted cells.

4.6.7 Future Work

The two Cre expressing lines will be crossed to mice that contain a single copy of the hCD2 transgene possessing a floxed HSS3 in a near centromeric position. This will allow the deletion of HSS3 in a tissue specific manner. In addition, a transgenic line that directs Cre recombinase expression later in development, after the initiation of hCD2 expression, still needs to be generated.

Chapter Five

Cre Recombinase Inducible Lineage Ablation

5.1 Experimental Rational

The previous section described the generation and testing of Cre recombinase expressing lines using the R26R-EYFP reporter line. However, activation of the R26R-EYFP reporter gene is a neutral event, i.e. the fate of the cells is not affected by the expression of EYFP. In such a system a low number of cells not expressing Cre would be hard to detect. Therefore, the Cre expressing lines were tested in a system that applied 'negative selection pressure' to cells that expressed Cre recombinase. This would give cells that do not express Cre a selective advantage, potentially allowing their detection more readily. To this end a system was generated that allowed Cre-inducible expression of the diphtheria toxin. Such mice would also be useful for conditional lineage specific ablation.

5.2 The Diphtheria Toxin

The diphtheria toxin (DT) is synthesised by *Corynebacterium diphtheriae* and is the causative agent of diphtheria [462]. The toxin is generated as a single polypeptide chain, which is secreted from the bacterium. After binding to the surface of the host cell the toxin peptide is cleaved by the protease furin, generating two fragments [463]. The N-terminal fragment, denoted the A-fragment (DT-A), has enzymatic activity and ADP-ribosylates the elongation factor E2F, resulting in cessation of protein synthesis and concomitant host cell death [464]. The C-terminal fragment, denoted the B-fragment (DT-B), is responsible for cell surface binding and translocation of the DT-A fragment into the cytosol [463]. Once in the cytosol, the DT-A fragment is extremely toxic and evidence suggests that a single molecule of this protein is sufficient for causing cell death [465].

The DT gene has become a useful tool in developmental biology, as expression of the DT-A fragment in a cell type specific manner allows the selective killing of cells within a defined lineage [466-472]. Even if released from the cytoplasm of dying cells the DT-A fragment is unable to enter surrounding cells in the absence of the DT-B fragment, thus preventing non-specific cell death [466]. For example, transgenic mice expressing an elastase-DT-A transgene show specific cell lineage ablation within the pancreas but not in any other tissue [466]. Similarly, transgenic mice expressing an H1tDTA transgene show specific ablation of male germ cells, but no other tissues are affected [473].

One major drawback of using a DT-A transgenic approach is that the mice frequently die prematurely or are infertile [472, 473]. As a result, establishing DT-A transgenic lines is problematic. To overcome this hurdle a number of Cre-inducible DT-A systems have been generated. In these systems a floxed transcriptional stop sequence is placed between the DT-A gene and the promoter, thereby preventing expression of the toxin [474-477]. This allows the generation and stable maintenance of DT-A transgenic mice. When these mice are bred to tissue-specific Cre expressing transgenic lines, recombinase mediated deletion of the floxed stop sequence induces expression of the DT-A gene resulting in lineage specific ablation.

In order to allow the regulated ablation of specific cell lineages a system was generated that allows the inducible expression of the DT-A gene in a temporally and spatially controlled manner. To this end, a Cre-loxP based approach was utilised in which expression of DT-A gene can be initiated by Cre mediate recombination. Previously a system has been developed which allows the targeting of any cDNA to the ubiquitously

expressed ROSA26 locus [446]. The system uses a generic targeting vector (pBigT), into which any cDNA can easily be inserted downstream of a floxed neo-tpA cassette, and then subsequently cloned into a plasmid containing ROSA26 genomic flanking arms. Following targeting of the ROSA26 locus the transcription of the cDNA is prevented by the presence of an upstream triple polyadenylation signal. Cre mediated recombination results in the excision of the stop signal initiating expression of the cDNA of interest.

5.3 Cloning

A vector (DT-A-pUC) containing a truncated DT gene encoding only for the A-fragment was obtained from Ian Maxwell. In this cassette the initial Met-Asp-Pro sequence is donated by the human metallothionein IIA sequence, followed by the DT-A sequence that encodes amino acids 3 through 193, followed by the SV40 sequence that provides a carboxyl-terminal Ser-Leu and the small t intron [466].

The DT-A gene was excised from the DT-A-pUC vector using BglII restriction sites flanking the gene. After blunting, this fragment was cloned into the EcoRV site of pBluescript (pBS) to give DT-pBS. This was done to generate a vector in which the DT-A gene was flanked by a number of useful restriction sites. The DT-A gene was lifted from this vector using a SalI-NotI restriction digest and cloned into SalI-NotI cut pBigT vector to give DT-pBigT. The DT-pBigT vector was cut with PacI-AscI and the fragment containing the floxed neo-tpA and DT-A gene was cloned into the ROSA26PA vector via PacI and AscI sites, to give the final DT-ROSA26PA targeting vector. Before targeting the DT-ROSA26PA vector was linearised with KpnI.

5.4 Targeting

The linearised construct was introduced into R1 ES cells by electroporation. $8x10^6$ cells were electroporated with 30μg of the linearised vector (see materials and methods). Selection was started 24 hours post electroporation by the addition of 300μg/ml of G418. A schematic diagram of the targeting is shown in figure 19A. Targeting of the ROSA26-1 locus with the DT-ROSA26PA vector results in the introduction of an additional EcoRV site into the genomic locus. Using an EcoRV digest and a 5' outside homology probe it is possible to differentiate between the wild type and targeted locus (figure 19B). Of the 72 clones that were analysed by Southern blot 4 were shown to have undergone homologous recombination. These 4 positive clones were selected for further Southern blot analysis. Insertion of the DT-ROSA26PA vector into the ROSA26-1 genomic locus also results in the introduction of an AscI site. Southern analysis using a PacI-AscI digest and a 3' outside homology probe allows the discrimination of the wild type and targeted locus. All of the analysed clones were confirmed to have undergone appropriate homologous recombination resulting in the insertion of the DT-ROSA26PA vector into the ROSA26-1 locus (data not shown). Clone 29 had a euploid karyotype (chromosome number of 40) and was selected to generate chimeric mice. ES cells were injected into 3.5 day-old blastocysts derived from C57BL/6 females and subsequently transferred into the uteri of foster mothers. Chimeric males were bred to C57BL/6 females and germline transmission was determined by agouti coat colour. Heterozygous mice carrying the targeted ROSA26-1 locus were identified by Southern blot analysis (figure 19C) and backcrossed to C57Bl/6 background. The resulting line was denoted DIP-R1. The observation that DIP-R1 transgenic mice are both viable and fertile indicates that the inducible system is not leaking (at least not to a detrimental level).

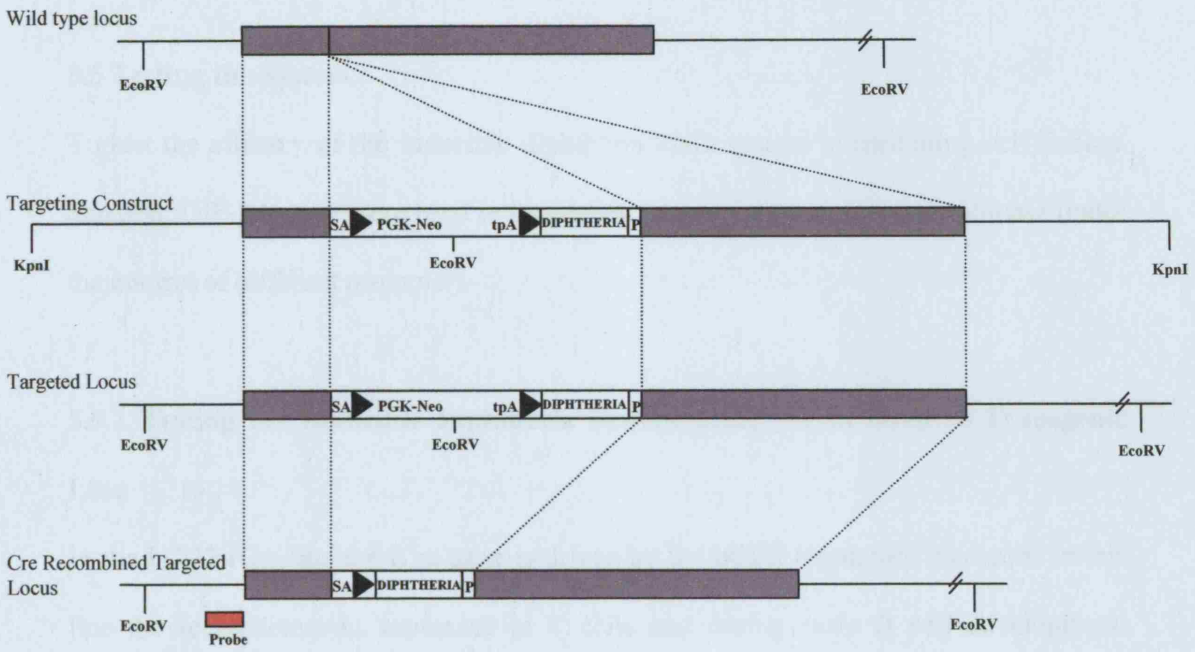
Figure 19. Targeting of the Diphtheria gene into the ROSA26 locus.

(A) Diagram of the wild type ROSA locus, targeting construct and targeted allele. The targeting construct contains the diphtheria gene down stream of a floxed neo gene and triple polyadenylation signal flanked by homologous DNA sequences from the ROSA locus (filled grey boxes). Black triangles denote loxP sites. Relevant restriction sites are indicated. The probe utilised in screening the ES clones is shown as a red bar. SA= splice acceptor. tpA= triple polyadenylation signal. P= polyadenylation signal.

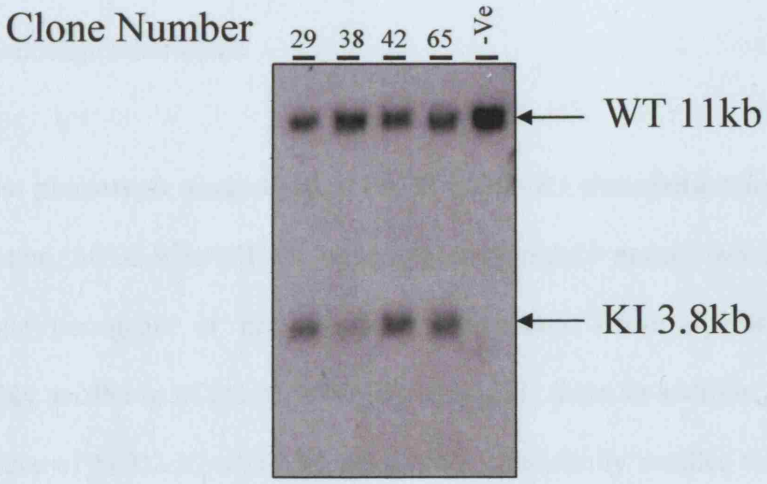
(B) Southern blot analysis of genomic DNA from 5 ES cell clones, digested with EcoRV and hybridized with the probe indicated in (A). The 11kb band is the wild type (WT) band and the 3.8kb band represents the targeted allele (KI).

(C) Southern blot analysis of DNA from DIP-R1 knockin mice digested with EcoRV and hybridized with the probe indicated in A. The 11kb band is the wild type (WT) band and the 3.8kb band represents the targeted allele (KI).

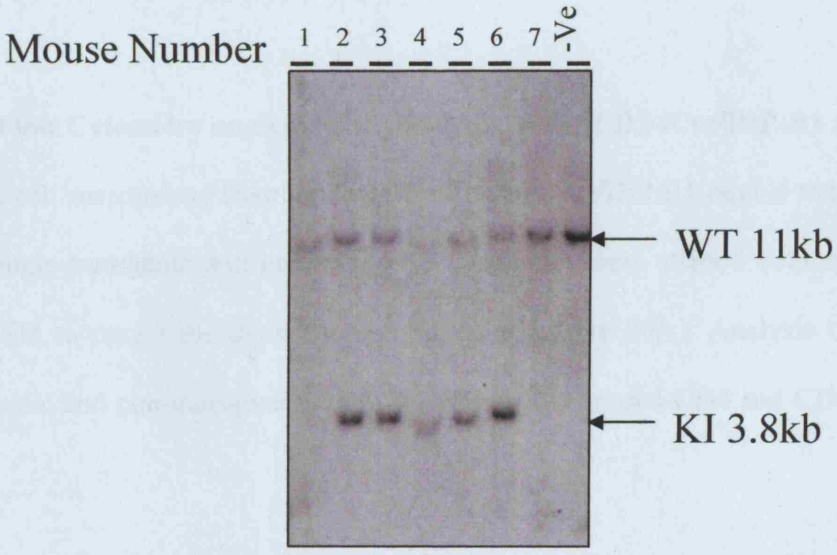
A.


Wild type locus
EcoRV
EcoRV
Targeting Construct
KpnI
SA
PGK-Neo
tpA
DIPHTHERIA
P
EcoRV
KpnI
Targeted Locus
EcoRV
SA
PGK-Neo
tpA
DIPHTHERIA
P
EcoRV
EcoRV
Cre Recombined Targeted Locus
EcoRV
Probe
SA
DIPHTHERIA
P
EcoRV

B.


Clone Number
29
38
42
65
-Ve
WT 11kb
KI 3.8kb

C.


Mouse Number
1
2
3
4
5
6
7
-Ve
WT 11kb
KI 3.8kb

5.5 Testing the System

To test the efficacy of the inducible diphtheria toxin system in mediating cell lineage ablation, DIP-R1 mice were bred to transgenic lines expressing Cre recombinase under the control of different promoters.

5.5.1 Testing the Inducible Diphtheria System using the hCD2-iCre Transgenic Line

In the hCD2-iCre line the iCre gene is driven by the hCD2 regulatory elements. In this line the recombinase is expressed in T cells and during early B cell development. Crossing this line to the DIP-R1 line is expected to result in mice in which both T cells and B cells are ablated. Double transgenic hCD2-iCre/DIP-R1 mice were born with normal Mendelian inheritance.

5.5.2 Gross phenotypic analysis of hCD2-iCre/DIP-R1 transgenic mice

On dissection, hCD2-iCre/DIP-R1 mice appeared grossly normal when compared to their single transgenic or non-transgenic littermates. However, the thymus was considerably smaller in hCD2-iCre/DIP-R1 transgenic mice. In addition, the spleen and lymph nodes of hCD2-iCre/DIP-R1 mice were consistently smaller than their single transgenic or non-transgenic littermates.

5.5.3 Flow Cytometry analysis of thymocytes from hCD2-iCre/DIP-R1 mice

Single cell suspensions from the thymus of hCD2-iCre/DIP-R1 double transgenic mice and single transgenic and non-transgenic littermates were stained with anti-CD4 and anti-CD8 to reveal the thymic sub-populations (figure 20A). Analysis of the single transgenic and non-transgenic littermates reveals the normal CD4 and CD8 expression

patterns expected in the thymus. Analysis of thymocytes from hCD2-iCre/DIP-R1 double transgenic mice shows that all subsets seen in single transgenic and non-transgenic littermates are present. However, it is apparent that in hCD2-iCre/DIP-R1 double transgenic mice there is a relative increase in the proportion of DN ($CD4^-/CD8^-$) cells.

Figure 20B shows the absolute numbers of DN and DP thymocytes in hCD2-iCre/DIP-R1 double transgenic mice as compared to single transgenic and wild type littermates. From this analysis it can be calculated that in hCD2-iCre/DIP-R1 double transgenic mice the DN thymocytes are decreased by approximately 50% whereas the DP thymocytes are decreased by more than 90% when compared to single transgenic and non-transgenic littermates. This differential decrease between the two populations can account for the apparent relative increase in the proportion of DN thymocytes over DP thymocytes in the hCD2-iCre/DIP-R1 double transgenic mice, as shown by flow cytometry.

5.5.4 Flow Cytometry Analysis of Splenocytes from hCD2-iCre/DIP-R1 Mice

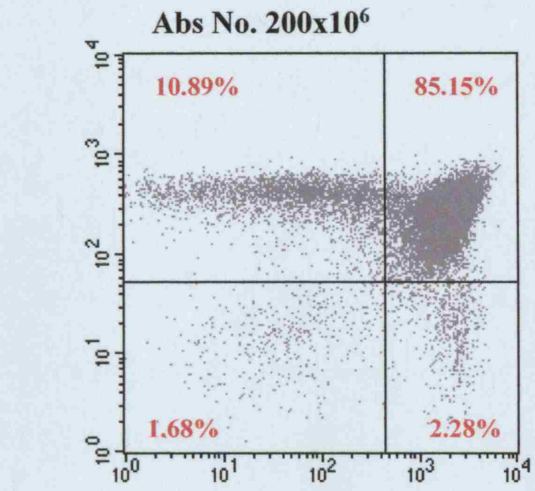
Single cell suspensions from the spleen of hCD2-iCre/DIP-R1 double transgenic mice, single transgenic and non-transgenic littermates were stained with anti-TCR (a cell surface T cell marker) and anti-CD19 (a cell surface B cell marker) to reveal the T cell and B cell populations. Analysis of splenocytes from single transgenic and non-transgenic littermates reveals the expected staining profile. Analysis of hCD2-iCre/DIP-R1 double transgenic mice revealed that the T cell and B cell populations within the spleen appear grossly normal (figure 21A). However, when the absolute numbers of T cells and B cells are calculated both are reduced in hCD2-iCre/DIP-R1 double

Figure 20. Flow cytometric analysis of thymocytes from hCD2-iCre/DIP-R1 mice.

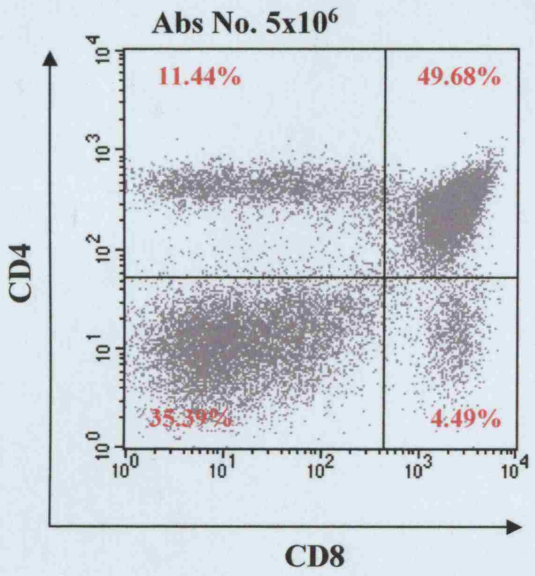
(A) Single cell suspensions from thymus were prepared from hCD2-iCre/DIP-R1 double transgenic mice and single transgenic or non-transgenic littermate controls. These cells were stained with CD4 and CD8 antibodies to allow differentiation of the stages of thymocyte development. The absolute number of thymocytes is indicated above the dot plots. The percentage of cells within each quadrant is indicated.

(B) Graphs showing the absolute numbers of DN and DP thymocytes in hCD2-iCre/DIP-R1 double transgenic mice and single transgenic or non-transgenic littermate controls. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated.

A.

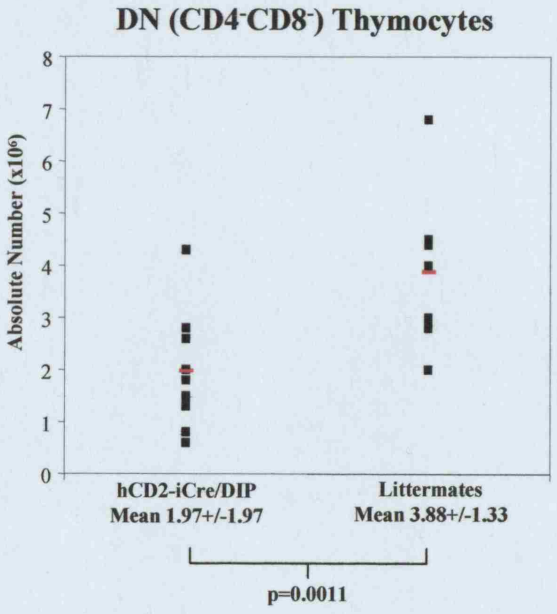


Littermate



hCD2-iCre/DIP-R1

B.



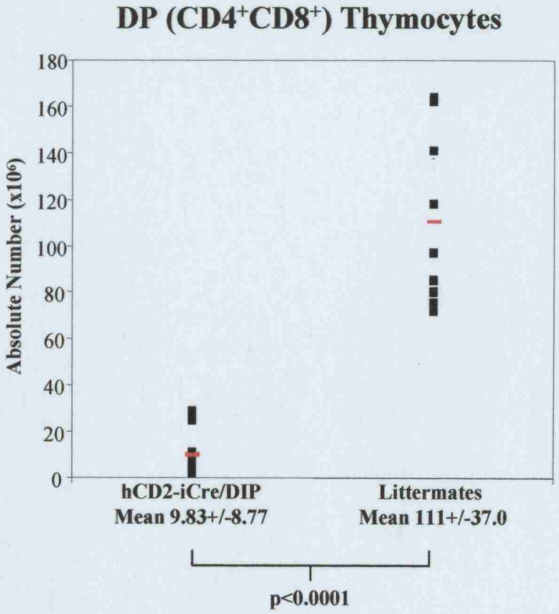
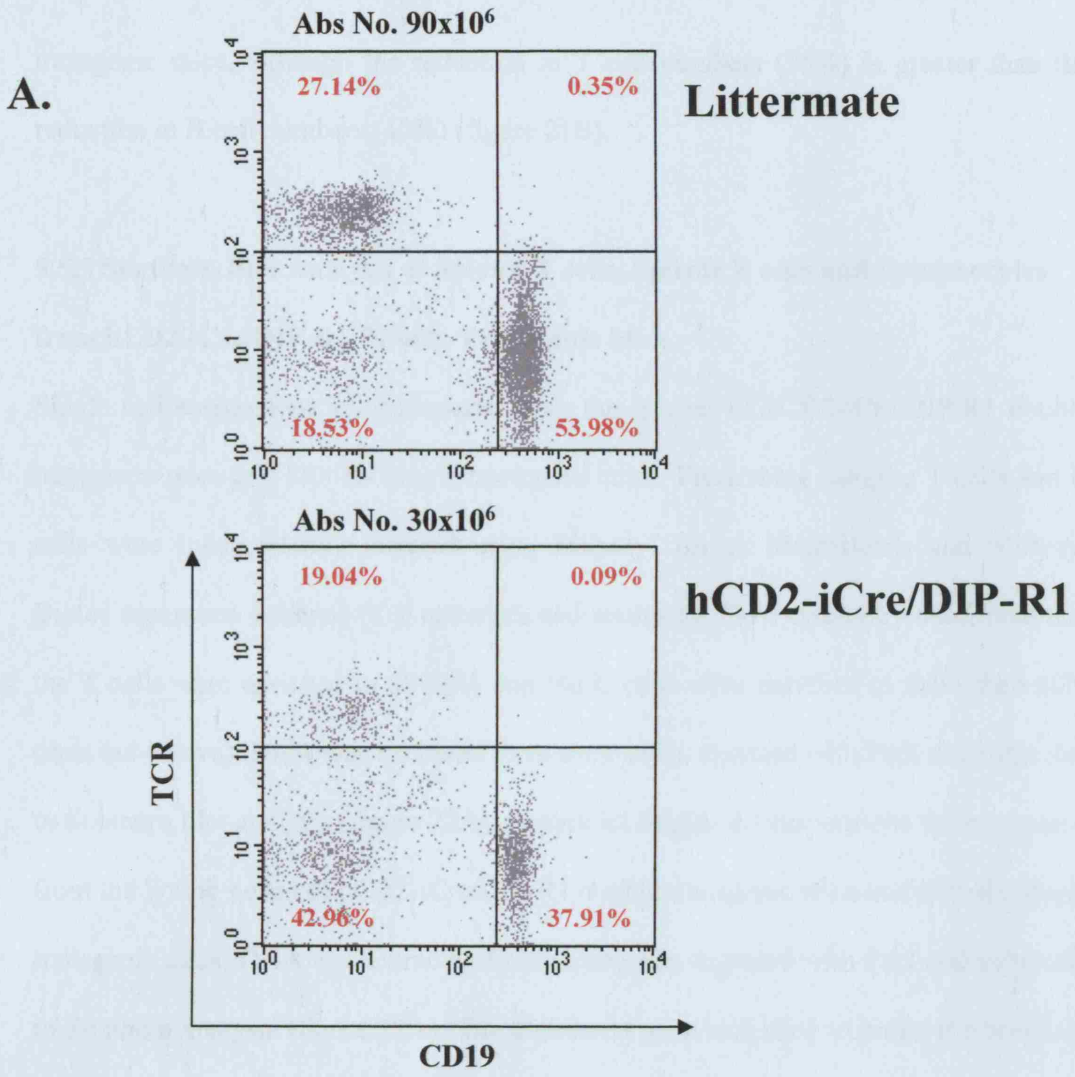
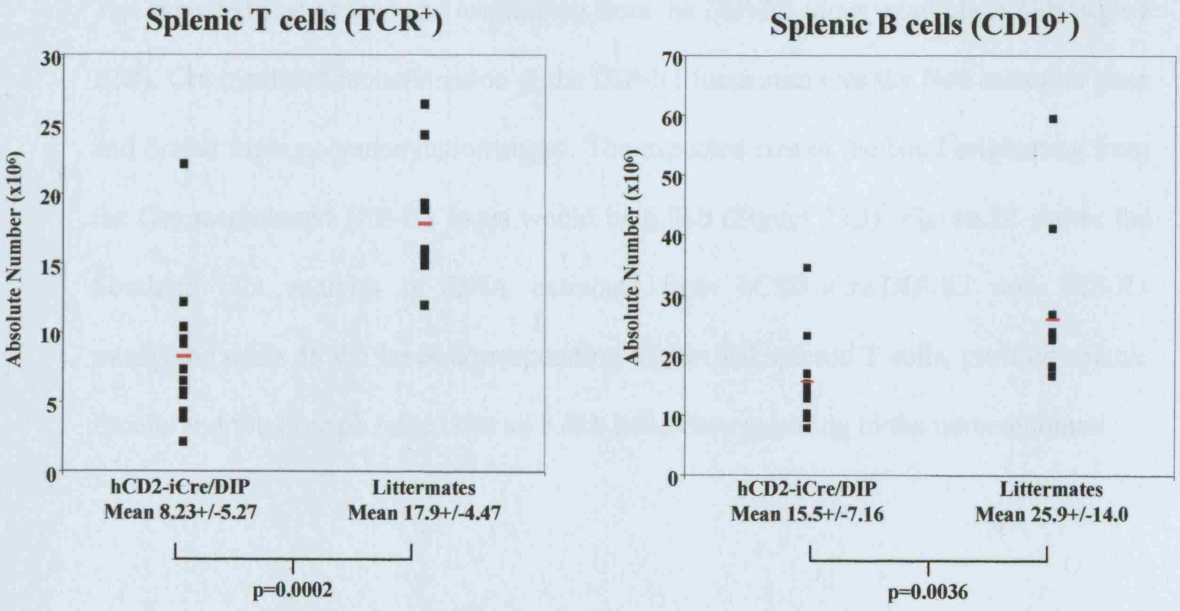


Figure 21. Flow cytometric analysis of splenocytes from hCD2-iCre/DIP-R1 mice.

(A) Single cell suspensions from spleen were prepared from hCD2-iCre/DIP-R1 double transgenic mice and single transgenic or non-transgenic littermate controls. These cells were stained with TCR and CD19 antibodies to allow differentiation of the T cell ($TCR^+/CD19^-$) and B cell populations ($CD3^-/CD19^+$). The absolute number of splenocytes is indicated above the dot plots. The percentage of cells within each quadrant is indicated.

(B) Graphs showing the absolute numbers of splenic T cells and B cells in hCD2-iCre/DIP-R1 double transgenic mice and single transgenic or non-transgenic littermate controls. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated.


A.
Abs No. 90x10⁶
27.14%
0.35%
18.53%
53.98%
Littermate
Abs No. 30x10⁶
19.04%
0.09%
42.96%
37.91%
hCD2-iCre/DIP-R1
TCR
CD19


B.
Splenic T cells (TCR⁺)
Absolute Number (x10⁶)
hCD2-iCre/DIP
Mean 8.23+/-5.27
Littermates
Mean 17.9+/-4.47
p=0.0002
Splenic B cells (CD19⁺)
Absolute Number (x10⁶)
hCD2-iCre/DIP
Mean 15.5+/-7.16
Littermates
Mean 25.9+/-14.0
p=0.0036

transgenic mice, although the reduction in T cell numbers (55%) is greater than the reduction in B cell numbers (40%) (figure 21B).

5.5.5 Southern Blot Analysis of Splenic T cells, Splenic B cells and Lymphocytes from hCD2-iCre/DIP-R1 Double Transgenic Mice

Single cell suspensions were prepared from the spleens of hCD2-iCre/DIP-R1 double transgenic mice and DIP-R1 single transgenic mice. From these samples T cells and B cells were independently isolated using Miltenyi Biotec MicroBeads and Miltenyi Biotec separation columns (See materials and methods). Flow cytometry confirmed that the T cells were enriched to 60-80% and the B cells were enriched to more than 80% (data not shown). DNA was extracted from these cells, digested with PacI and subjected to Southern blot analysis (figure 22A). In parallel single cell suspensions were prepared from the lymph nodes of hCD2-iCre/DIP-R1 double transgenic mice and DIP-R1 single transgenic mice. DNA was extracted from these cells, digested with PacI and subjected to Southern analysis (figure 22A). The diphtheria gene was used to probe the Southern blot.

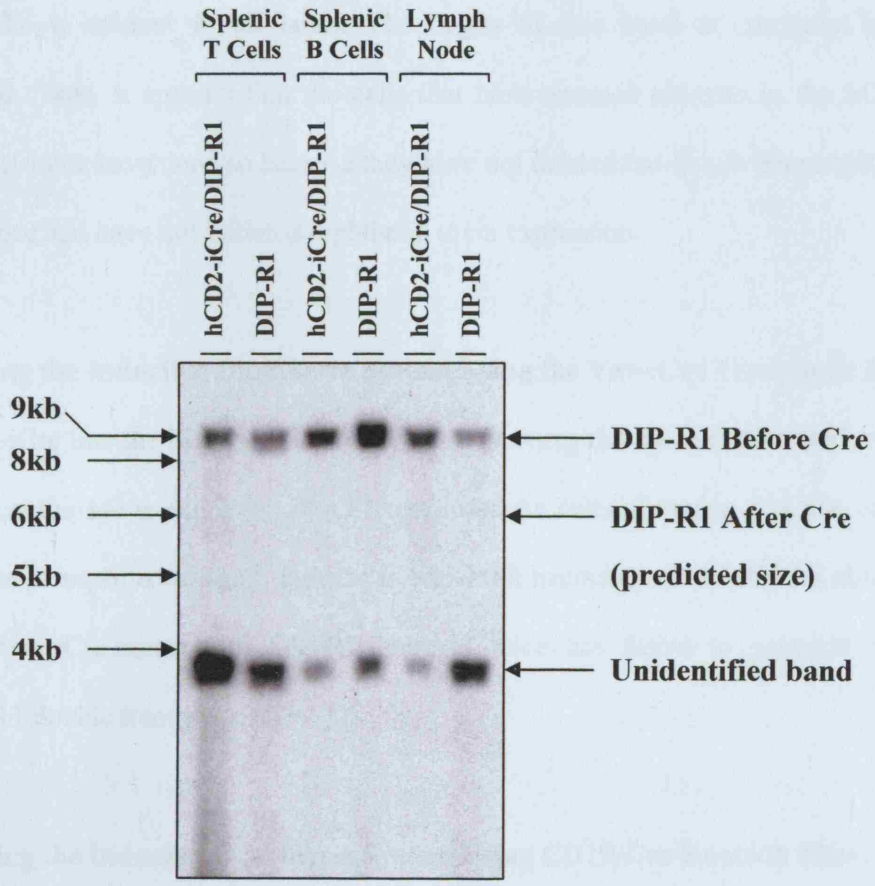
The expected size of the band originating from the DIP-R1 locus would be 8.6kb (figure 22B). Cre mediated recombination at the DIP-R1 locus removes the Neo selection gene and floxed triple polyadenylation signal. The expected size of the band originating from the Cre recombined DIP-R1 locus would be 6.0kb (Figure 22B). Figure 22 shows the Southern blot analysis of DNA extracted from hCD2-iCre/DIP-R1 and DIP-R1 transgenic mice. In the lanes corresponding to purified splenic T cells, purified splenic B cells and total lymph node cells an 8.6kb band corresponding to the unrecombined

Figure 22. Southern blot analysis of splenic T cells, splenic B cells and lymph node cells from hCD2-iCre/DIP-R1 double transgenic mice and DIP-R1 single transgenic mice.

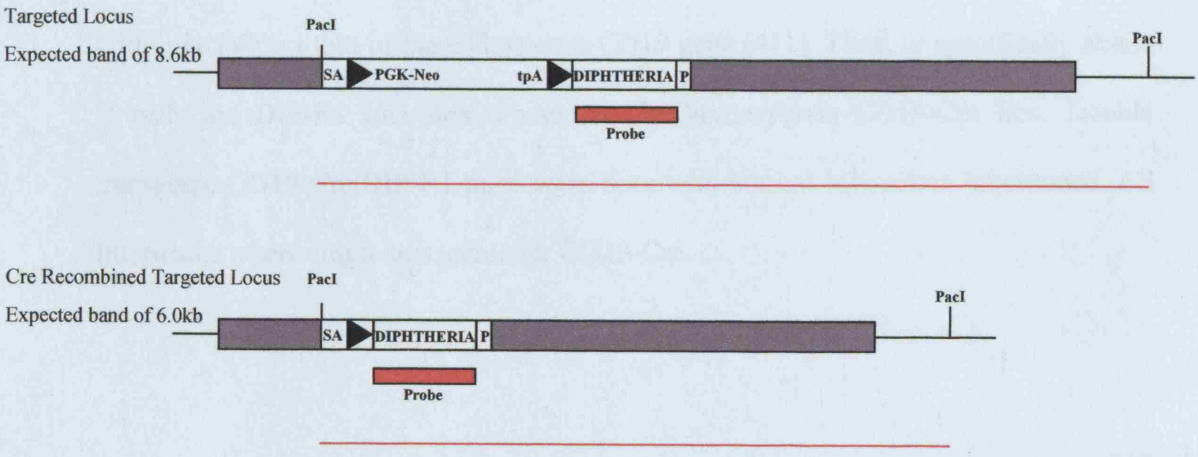
(A) T cells and B cells were independently purified from the spleens of hCD2-iCre/DIP-R1 double transgenic mice and DIP-R1 single transgenic mice. DNA was extracted from these cells, digested with PacI and subjected to Southern blot analysis. In parallel single cell suspensions were prepared from the lymph nodes of hCD2-iCre/DIP-R1 double transgenic mice and DIP-R1 single transgenic mice. DNA was extracted from these cells, digested with PacI and subjected to Southern analysis. The diphtheria gene was used to probe the Southern blots.

(B) Schematic diagram of the DIP-R1 knockin locus before and after Cre mediated recombination. Black triangles denote loxP sites. The probe utilised is shown as a red bar. The fragments detected with the diphtheria probe are represented by a thin red line. SA= splice acceptor. tpA= triple polyadenylation signal. P= polyadenylation signal.

A.


Splenic T Cells
Splenic B Cells
Lymph Node
hCD2-iCre/DIP-R1
DIP-R1
hCD2-iCre/DIP-R1
DIP-R1
hCD2-iCre/DIP-R1
DIP-R1
9kb
8kb
6kb
5kb
4kb
DIP-R1 Before Cre
DIP-R1 After Cre
(predicted size)
Unidentified band

B.


Targeted Locus
Expected band of 8.6kb
PacI
SA
PGK-Neo
tpA
DIPHTHERIA
P
PacI
Probe
Cre Recombined Targeted Locus
Expected band of 6.0kb
PacI
SA
DIPHTHERIA
P
PacI
Probe

DIP-R1 locus is detected. In contrast, there is no evidence for a 6kb band (that would correspond to the Cre recombined DIP-R1 locus) in any of the samples. An additional band of 4kb is evident in all lanes. The origin of this band is currently being investigated. Thus, it appears that the cells that have escaped ablation in the hCD2-iCre/DIP-R1 mice have done so because they have not deleted the floxed transcriptional stop sequence and have not initiated diphtheria toxin expression.

5.5.6 Testing the Inducible Diphtheria System Using the Vav-iCre Transgenic Line

In the Vav-iCre line the iCre gene is driven by the Vav regulatory elements. In this line the Cre recombinase is expressed in all haematopoietic cells. Crossing this line to the DIP-R1 line is expected to result in mice in which all haematopoietic cells are ablated. Crossing Vav-iCre mice with DIP-R1 knockin mice has failed to generate Vav-iCre/DIP-R1 double transgenic mice.

5.5.7 Testing the Inducible Diphtheria System Using CD19-Cre Knockin Mice

In the CD19-Cre line, the Cre recombinase gene is knocked into the endogenous CD19 locus in a manner that utilises the CD19 regulatory elements to drive expression [411]. In wild type mice, CD19 expression is initiated in Pro-B cells and is maintained throughout development to the mature B cell stage where its expression is upregulated [478]. Cre recombinase expression in the CD19-Cre line is strictly regulated and faithfully follows that of the endogenous CD19 gene [411]. Thus, to specifically ablate B cells the DIP-R1 line was crossed to the homozygous CD19-Cre line. Double transgenic CD19-Cre/DIP-R1 mice were born with normal Mendelian inheritance. All littermates where single transgenic for CD19-Cre.

5.5.8 Gross Phenotypic Analysis of DIP-R1/CD19-Cre double transgenic mice

On dissection, CD19-Cre/DIP-R1 mice appeared grossly normal when compared to their single transgenic littermates. However, the spleen from CD19-Cre/DIP-R1 mice was consistently smaller than from single transgenic littermates. The thymus and lymph nodes appeared normal in size and gross morphology.

5.5.9 Flow Cytometry of Splenocytes from CD19-Cre/DIP-R1 mice

Single cell suspensions from the spleen of CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates were stained with anti-TCR and anti-CD19 to reveal the T cell and B cell populations. Analysis of splenocytes from CD19-Cre single transgenic littermates reveals the expected staining profile. Analysis of CD19-Cre/DIP-R1 double transgenic mice revealed a decrease in the proportion of B cells within the spleen (figure 23A). Calculating the total numbers of B cells within the spleen of each mouse revealed that the absolute splenic B cell number is reduced by approximately 60% when compared to CD19-Cre single transgenic littermate controls (figure 23B). There is no significant difference in the absolute number of splenic T cells between CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates (data not shown).

5.5.10 Flow Cytometry of Lymph Node Derived Lymphocytes from CD19-Cre/DIP-R1 Mice

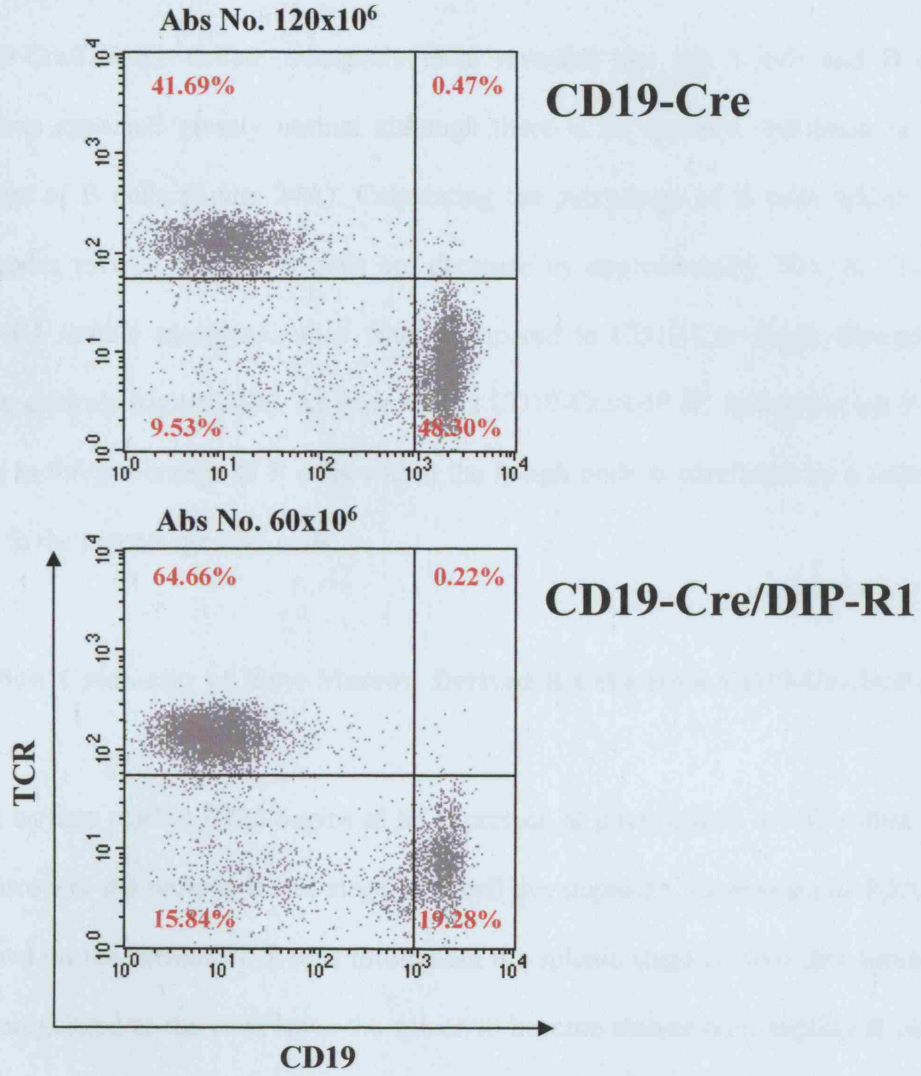
Single cell suspensions from the lymph nodes of CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates were stained with anti-TCR and anti-CD19 to reveal the T cell and B cell populations. Analysis of lymph node cells from CD19-Cre single transgenic littermates reveals the expected staining profile. Analysis

Figure 23. Flow cytometric analysis of splenocytes from CD19-Cre/DIP-R1 mice.

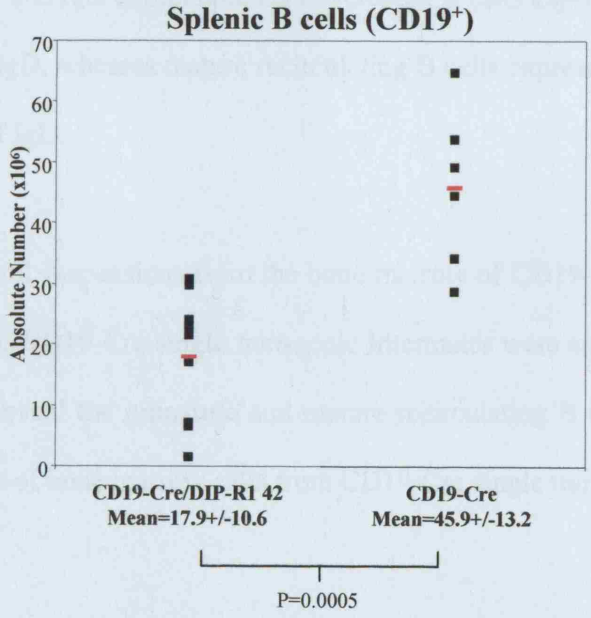
(A) Single cell suspensions from spleen were prepared from CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermate controls. These cells were stained with TCR and CD19 antibodies to allow differentiation of the T cell ($TCR^+/CD19^-$) and B cell populations ($CD3^-/CD19^+$). The absolute number of splenocytes is indicated above the dot plots. The percentage of cells within each quadrant is indicated.

(B) Graph showing the absolute numbers of splenic B cells in CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermate controls. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated.

A.


Abs No. 120x10⁶
41.69%
0.47%
9.53%
48.30%
CD19-Cre
Abs No. 60x10⁶
64.66%
0.22%
15.84%
19.28%
CD19-Cre/DIP-R1
TCR
CD19

B.


Splenic B cells (CD19⁺)
Absolute Number (x10⁶)
CD19-Cre/DIP-R1 42
Mean=17.9+/-10.6
CD19-Cre
Mean=45.9+/-13.2
P=0.0005

of CD19-Cre/DIP-R1 double transgenic mice revealed that the T cell and B cell populations appeared grossly normal although there is an apparent reduction in the percentage of B cells (figure 24A). Calculating the percentage of B cells within the lymph nodes reveals that the B cells are decrease by approximately 70% in CD19-Cre/DIP-R1 double transgenic mice when compared to CD19-Cre single transgenic littermate controls (figure 24B). As expected, in CD19-Cre/DIP-R1 transgenic mice the decrease in the percentage of B cells within the lymph node is paralleled by a relative increase in the percentage of T cells.

5.5.11 Flow Cytometry of Bone Marrow Derived B Cells from CD19-Cre/DIP-R1 Mice

The cell surface marker B220 begins to be expressed at intermediate levels within the bone marrow at the pre-pro-B cell stage of B cell development. Expression of B220 is maintained on the surface of B cells throughout the splenic stage of their development and is upregulated as the cells leave the spleen to become mature recirculating B cells. Mature recirculating B cells also express high levels of the cell surface marker IgD. Within the bone marrow two different populations can therefore be defined on the basis of B220 and IgD expression, i.e. developing B cells express intermediate levels of B220 and no IgD, whereas mature recirculating B cells express high levels of B220 and high levels of IgD.

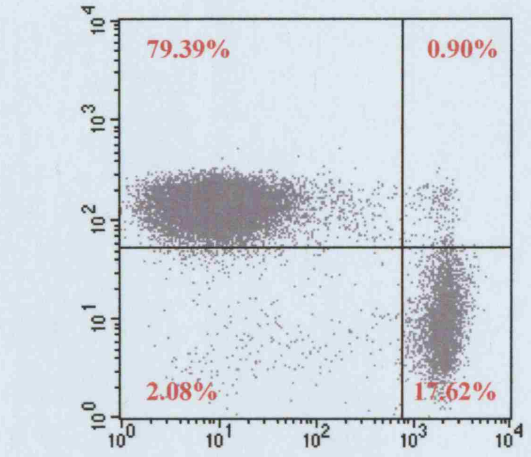
Single cell suspensions from the bone marrow of CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates were stained with anti-B220 and anti-IgD to reveal the immature and mature recirculating B cell populations (Figure 25A). Analysis of bone marrow cells from CD19-Cre single transgenic littermates reveals the

Figure 24. Flow cytometric analysis of lymph node derived lymphocytes from CD19-Cre/DIP-R1 mice.

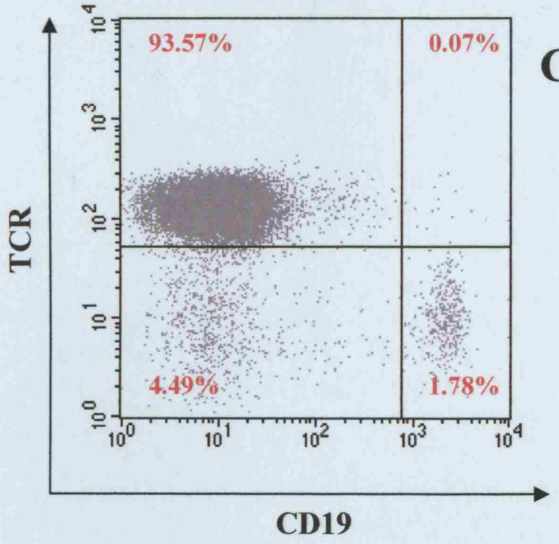
(A) Single cell suspensions from lymph node were prepared from CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermate controls. These cells were stained with TCR and CD19 antibodies to allow differentiation of the T cell ($TCR^+/CD19^-$) and B cell populations ($CD3^-/CD19^+$). The percentage of cells within each quadrant is indicated.

(B) Graph showing the percentage of lymph node B cells in CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermate controls. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated.

A.



CD19-Cre



CD19-Cre/DIP-R1

B.

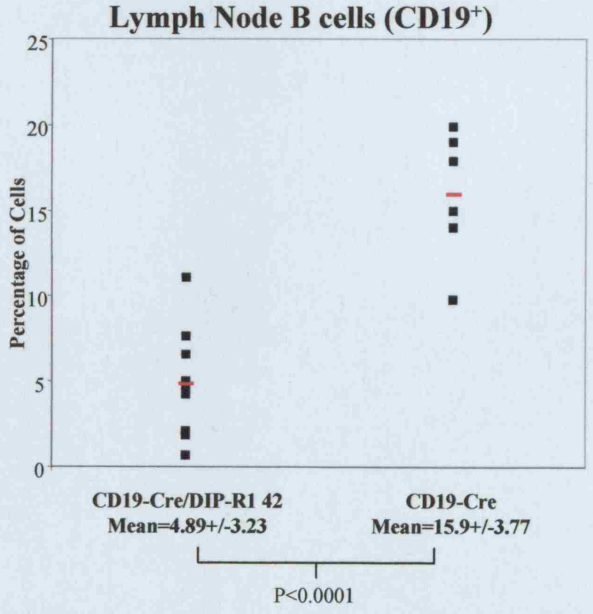
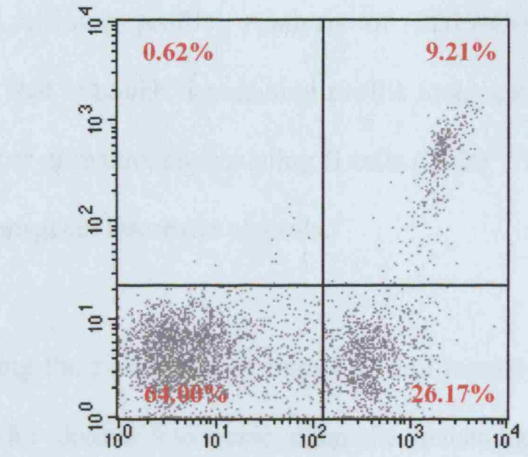


Figure 25. Flow cytometric analysis of bone marrow derived B cells from CD19-Cre/DIP-R1 transgenic mice.

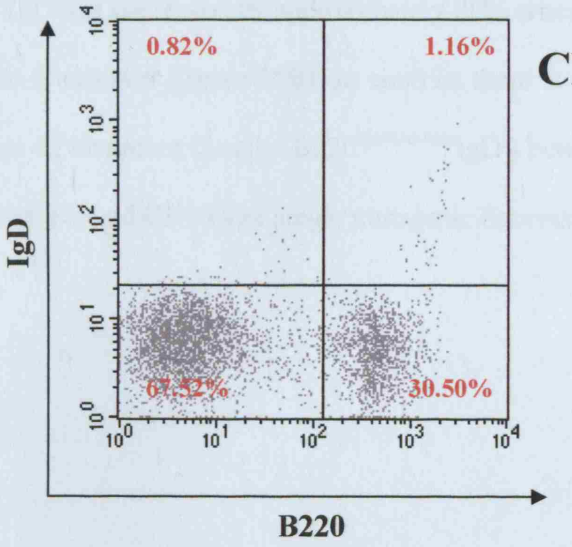
(A) Single cell suspensions from bone marrow were prepared from CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates. These cells were stained with IgD and B220 antibodies to allow differentiation of the immature (IgD^- /$B220^{intermediate}$) and mature recirculating B cells populations (IgD^+/$B220^{high}$). The percentage of cells within each quadrant is indicated.

(B) Graph showing the percentage of mature recirculating B cells in CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermate controls. Mean, standard deviation and statistical analysis (Student's T Test) of each data set is indicated.

A.

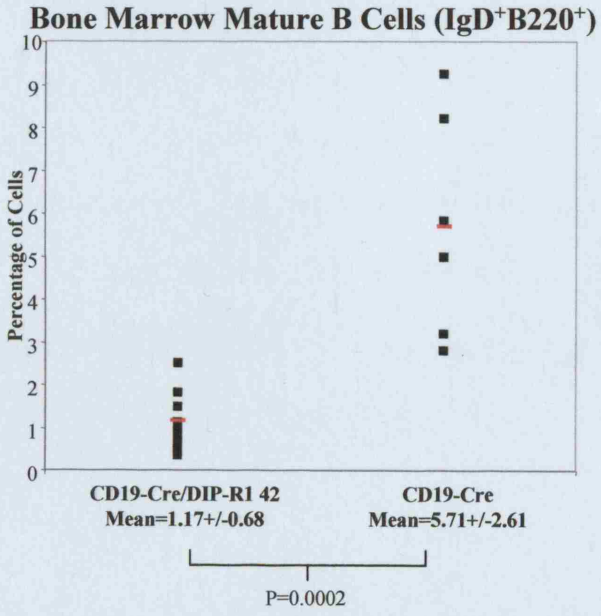


CD19-Cre



CD19-Cre/DIP-R1

B.



expected staining profile. Analysis of CD19-Cre/DIP-R1 double transgenic mice revealed that although the staining profile looks grossly normal there is a reduction in the number of mature recirculating B cells ($B220^{high}IgD^{+}$) when compared to CD19-Cre single transgenic littermate controls.

Calculating the percentage of immature and mature B cells reveals that within CD19-Cre/DIP-R1 double transgenic mice the percentage of mature recirculating B cells ($B220^{high}IgD^{+}$) is decreased by approximately 80% when compared to CD19-Cre single transgenic littermates (figure 25B). In contrast there is no significant difference in the percentage of immature B cells ($B220^{intermediate}IgD^{-}$) between CD19-Cre/DIP-R1 double transgenic mice and CD19-Cre single transgenic littermates (data not shown).

5.6 Discussion

We have developed a cell lineage ablation system that allows Cre-mediated inducible expression of the DT-A gene in a temporally and spatially controlled manner. To test the system, inducible diphtheria toxin mice were bred to the hCD2-iCre and CD19-Cre transgenic lines. In both cases Cre expression resulted in the lineage specific ablation. However, ablation was never 100% efficient. In contrast, breeding has failed to generate any Vav-iCre/DIP-R1 double transgenic mice. This indicates that the presence of the Vav-iCre and DIP-R1 genes may be lethal during embryogenesis.

5.6.1 Analysis of hCD2-iCre/DIP-R1 Transgenic Mice

The reduced thymic cellularity noted in hCD2-iCre/DIP-R1 transgenic mice reveals that the system is able to effectively drive lineage specific ablation. However, even with a >90% reduction in thymic cellularity (as noted in some mice) 5-10 million thymocytes still remain. This is a significant number of cells when the proliferative capacity of thymocytes is taken into account. Indeed, flow cytometry revealed that the remaining cells were able to complete their developmental program to become SP T cells and be exported to the periphery.

In the hCD2-iCre line initiation of Cre expression begins early in DN thymocyte development and increase until virtually 100% of DP thymocytes have expressed or are expressing Cre. Thus, within hCD2-iCre/DIP-R1 mice ablation is expected to initiate within the DN compartment and continue as cells develop into DP thymocytes. This is reflected by the lower level of ablation in the DN compartment than in the DP compartment.

Analysis of peripheral lymphocytes revealed a reduction in both the T cell and B cell compartments. However, the reduction in splenic cellularity (as well as absolute T cell numbers) is far less than the decrease in thymic cellularity. This suggests that cells that have escaped deletion in the thymus are able to expand in the periphery. However, the reduced splenic cellularity indicates that a large proportion of the cells that have escaped to the periphery are also ablated. Interestingly, the reduction in peripheral B cells is lower than that seen in the T cell compartment. Experiments utilising hCD2-iCre/R26R-EYFP transgenic mice indicates that Cre is expressed in both B cell and T cell progenitors (see Fig), whereas experiments utilising hCD2-GFP transgenic mice reveal that expression from the hCD2 regulatory elements is high in immature bone marrow B cells but shuts down in mature B cells. This could explain the decreased affect on the mature B cell pool, i.e. when a B cell reaches the periphery it has a decreased chance of activating the diphtheria toxin as expression of Cre has ceased in the majority of cells. In contrast, as the hCD2-iCre construct should always be expressed in mature T cells there is a greater chance of activating the diphtheria toxin in these cells.

Southern blot analysis of hCD2-iCre/DIP-R1 transgenic mice revealed that, in peripheral lymphocytes that had escaped ablation, the DIP-R1 locus had not undergone Cre mediated recombination, and as a result these cells had not expressed the diphtheria toxin. This implies that these cells have escaped ablation due to a lack of Cre expression or Cre mediated recombination at the DIP-R1 locus, and not due to a lack of diphtheria toxin toxicity.

5.6.2 Analysis of CD19-Cre/DIP-R1 Transgenic Mice

The reduction in the number of B Cells in CD19-Cre/DIP-R1 mice confirmed that the inducible diphtheria toxin is able to mediate lineage specific ablation. However, as noted in the hCD2-iCre/DIP-R1 mice, deletion was not 100% efficient. For example, there was a 60% reduction in splenic B cells, a 70% reduction in lymph node T cells and an 80% decrease in mature recirculating B cells within the bone marrow of hCD2-iCre/DIP-R1 mice. Interestingly, the different level of B cell deletion in each of these tissues suggests that as the B cells mature, from splenic B cells to lymph node B cells and finally to mature recirculating B cells within the bone marrow, the level of deletion progressively increases. This finding fits with the previously published observation that in CD19-Cre mice the level of Cre protein increases with progression of B cell development in parallel with endogenous CD19 expression [411]. Consequently, deletion by CD19-Cre is an ongoing process throughout B cell lymphopoiesis, which results in lower deletion efficiency in bone marrow pre-B cells compared to mature B cells in the spleen [411]. It has been estimated that the efficiency of deleting a floxed allele is <80% in bone marrow pre-B cells but increases to >90% in mature splenic B cells [411]. Importantly, however, these experiments were done using a floxed allele that is functionally neutral in B cells. In contrast, two publications have reported experiments in which the CD19-Cre mouse line has been used to mediate recombination events leading to decreased B cell survival. In both cases Cre mediated recombination resulted in a 2-4 fold reduction in absolute B cell numbers [479, 480]. Interestingly, however, Southern analysis revealed that in the remaining B cell populations only 40-70% of the B cells had undergone Cre mediated recombination [480]. These published results show that in the presence of negative selection pressure a significant proportion of cells are able to escape Cre mediated recombination. The recombination efficiencies

reported in these two independent publications are compatible with the levels of B cells escaping ablation reported herein.

5.6.3 How Do These Cells Escape Deletion?

The presence of T cells and B cells in the hCD2-iCre/DIP-R1 mice and the presence of B cells in the CD19-Cre/DIP-R1 mice indicate that significant proportion of cells had escaped ablation in both lines. There are three possible mechanisms to explain how these cells escape deletion. First, it is possible that all cells express Cre and have induced expression of the diphtheria toxin but some cells are able to survive the toxic effects. Early estimates suggested that a single molecule is sufficient to kill a cell [465]. In contrast, a subsequent study has reported that 300-30,000 molecules are required to kill a cell [466]. However, it must be noted that these later experiments were performed using fertilised mouse oocytes, which are several hundred times larger than somatic cells [466]. Second, Cre may not be expressed in all cells allowing a small proportion to not activate the diphtheria gene. Third, it is possible that, although Cre had been expressed, recombination had not occurred at the ROSA26-DIP-R1 locus, resulting in a failure to express the toxin. Data obtained from hCD2-iCre/R26R-EYFP mice does not provide any support to suggest that either of these two mechanisms are probable events. However, Southern blot analysis reveals that the cells escaping ablation in hCD2-iCre/DIP-R1 transgenic mice had not activated the diphtheria gene. Thus, it appears the failure to ablate all peripheral lymphocytes in hCD2-iCre/DIP-R1 transgenic mice results from inefficiencies in either Cre expression or Cre mediated recombination, and not from a lack of diphtheria toxin toxicity.

In addition, a similar inducible diphtheria system has recently been described. Brockschnieder et al. knocked the lacZ gene and diphtheria toxin gene into the ROSA locus [475]. The system was tested by crossing the inducible diphtheria mice to different Cre expressing lines [475]. Interestingly, when bred to the CD19-Cre line only a 50% reduction in B cells was detected [475]. Using Southern analysis they determined that B cells which had escaped deletion had not undergone Cre mediated recombination [475]. Furthermore, analysis of splenocytes revealed a 25% increase in the level of proliferation, suggesting that B cells which had escaped deletion expanded to partially replenish the B cell compartment [475]. Thus, it appears that the ability of B cells to escape deletion stems from an inefficiency inherent to the CD19-Cre line (as mentioned above), rather than a lack of diphtheria toxicity. It is therefore interesting to note that in the CD19-Cre transgenic line the neomycin selection marker utilised during targeting of the Cre recombinase into the CD19 locus has not been removed [411]. Selection markers that are maintained within a targeted locus have been shown to be able to disrupt the expression of the targeted and neighbouring genes [481]. Thus, the selection marker within the CD19-Cre may cause silencing of Cre expression in a low percentage of cells. This possibility is supported by the contrasting results obtained by crossing the inducible diphtheria mice to the CNP-Cre strain (that expresses Cre recombinase in oligodendrocyte precursors), which resulted in complete ablation of oligodendrocytes, thereby confirming the applicability of the inducible diphtheria toxin system [475].

For each of the three mechanisms described above, cells that have escaped deletion would have a selective advantage. This is compounded by the fact that the immune system is an extremely adaptive 'organ' and has an amazing regenerative capacity [482]. For example, the transfer of a small number ($1\text{-}2x10^6$) of un-purified bone

marrow cells into lethally irradiated mice (which lack all T cells, B cells and haematopoietic stem cells) results in the re-establishment of a complete immune system within ~6-12 weeks [482]. Furthermore, when low numbers of T cells are transferred into an 'empty host' (i.e. a mouse lacking T cells) they rapidly proliferate to partially replenish the T cell compartment [483]. Similarly, when mature B cells are transferred into an 'empty host' (i.e. a mouse lacking B cells) a limited number migrate to the spleen where they undergo an expansion of >30 fold to form a self-renewing persistent population [484]. However, neither T cells or B cells have the expansion potential to entirely replenish the peripheral compartment, requiring a continuous low level of T cell production from the thymus or B cell production from the bone marrow [485, 486].

As a result of this homeostatic expansion, even a low percentage of cells escaping deletion would be able to expand and constitute an almost complete immune system. Testing the DIP-R1 mice using a transgenic line that expresses Cre in a less adaptive tissue may reveal the true effectiveness of the system. In addition, using mice that are homozygous for both the Cre recombinase and diphtheria toxin may potentially reduce the percentage of cells that are able to escape deletion.

5.6.4 Summary

We have developed a system for Cre-mediated cell linage ablation. Testing the inducible diphtheria toxin using the hCD2-iCre and CD19-Cre recombinase expressing lines revealed that using this system it is possible to selectively kill cells of specific lineages. However, the system is not 100% efficient resulting in a significant proportion of cells escaping ablation. In hCD2-iCre/DIP-R1 transgenic mice the mechanism responsible for this inefficiency resulted from a failure of Cre recombinase to induce

expression of the diphtheria toxin and not a lack of diphtheria toxin toxicity. In conclusion, this system potentially provides a useful tool to address the function of specific cell types during development and also to determine the deletion efficiency of Cre expressing lines.

Section Two

Identification of Additional Regulatory Elements within the hCD2 Gene

Experimental Rational

Most research on the hCD2 gene has focused on the LCR and more specifically on the function of HSS3 and transcription factors binding within this sequence. Therefore, using a number of approaches, including bioinformatics, additional mechanisms potentially responsible for the regulation of hCD2 expression were investigated.

Chapter Six

Analysis of Potential Matrix Attachment Regions

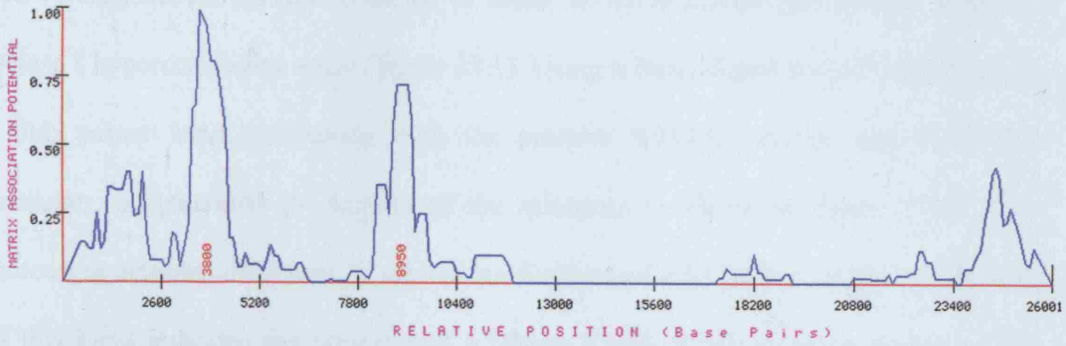
6.1 Scaffold/Matrix Attachment Regions

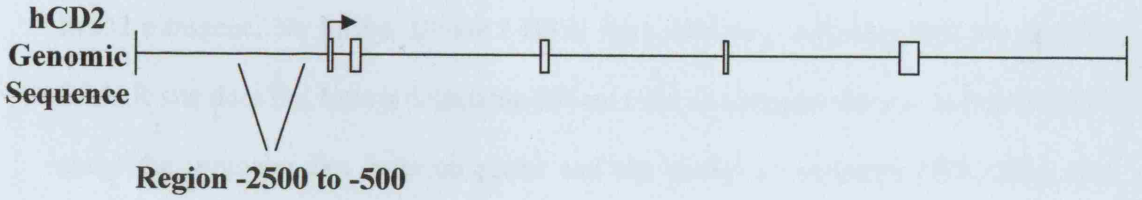
In the radial loop model of chromosome structure AT-rich DNA sequences found at the base of each chromatin loop are thought to mediate attachment to a proteinaceous scaffold [182]. These sequences have been termed scaffold or matrix attachment regions (S/MARs) and are thought to have important roles in maintaining chromatin structure and regulating gene expression. [218]. A typical S/MAR is comprised of a number of key nucleotide motifs, none of which are highly specific. Most obviously, binding of S/MARs to the nuclear matrix correlates with AT-rich motifs, which are thought to function as DNA-unwinding elements. [218]. S/MARs are also enriched in potential replication origins, homeotic protein-binding sites, topoisomerase binding and cleavage sites, intrinsically curved or kinked DNA as well as S/MAR recognition sequences [218]. Based on computational prediction analysis one S/MAR site is expected to be found every 9-10kb in both the mouse and the human genomes, which is consistent with the spacing of experimentally defined S/MARs reported for several human gene clusters [218].

Previously S/MAR elements have been shown to protect against position effects and to function as boundary elements in transgenic mice [212, 487, 488]. S/MAR sites have also been associated with LCR sequences [388, 489, 490]. Therefore it was interesting to identify potential S/MAR sites within the hCD2 gene. A number of computational tools are available for the prediction of S/MAR sites based on the frequency of S/MAR–motifs within a query sequence. The MAR-Wiz computational software was used to screen the hCD2 genomic sequence for the potential presence of S/MAR sites (figure 26). The search identified a S/MAR site with high potential immediately 5' of the

Figure 26. MAR-Wiz analysis of the hCD2 genomic sequence.

The structure of the hCD2 genomic locus is shown below a graphical representation of potential MAR sites within the hCD2 gene. The approximate region containing the predicted MAR site is shown (-2500 to -500 with respect to the transcriptional start site). Open boxes represent exons.


1.00
0.75
0.50
0.25
MATRIX ASSOCIATION POTENTIAL
3800
8950
2600
5200
7800
10400
13000
15600
18200
20800
23400
26001
RELATIVE POSITION (Base Pairs)


hCD2
Genomic
Sequence
Region -2500 to -500

promoter within a region spanning the sequence approximately -2500 to -500 bp, with respect to the initiation site.

6.2 DNase I HSS Analysis of hCD2 Transgenic Mice

Previously, in human cell lines, a DNase I HSS has been identified within the region spanning the putative S/MAR site [491]. Thus, to determine whether this region is able to form a DNase I HSS in hCD2 transgenic mice, thymocytes from an homozygous MG4 transgenic mouse (that contains 12 copies of the hCD2 minigene) were used in a DNase I hypersensitivity assay (figure 27A). Using a BglII digest and a 5'cDNA probe a 3kb parent band, containing both the putative S/MAR element and the hCD2 promoter, is generated (a diagram of the minigene is shown in figure 27B). With increasing amounts of DNase I a second band of around 0.5-0.6kb is produced. The size of this band indicates the presence of a DNase I HSS at the promoter region of the hCD2 transgene. No further DNase I HSSs were detected, indicating that the putative S/MAR site does not form a detectable DNase I site in transgenic mice. In this DNase I assay the promoter lies between probe and the predicted upstream HSS. Thus, the strong promoter DNase I HSS may mask the presence of the upstream HSS if it is considerably weaker. The DNase I analysis was repeated using an alternative restriction digest and probe.

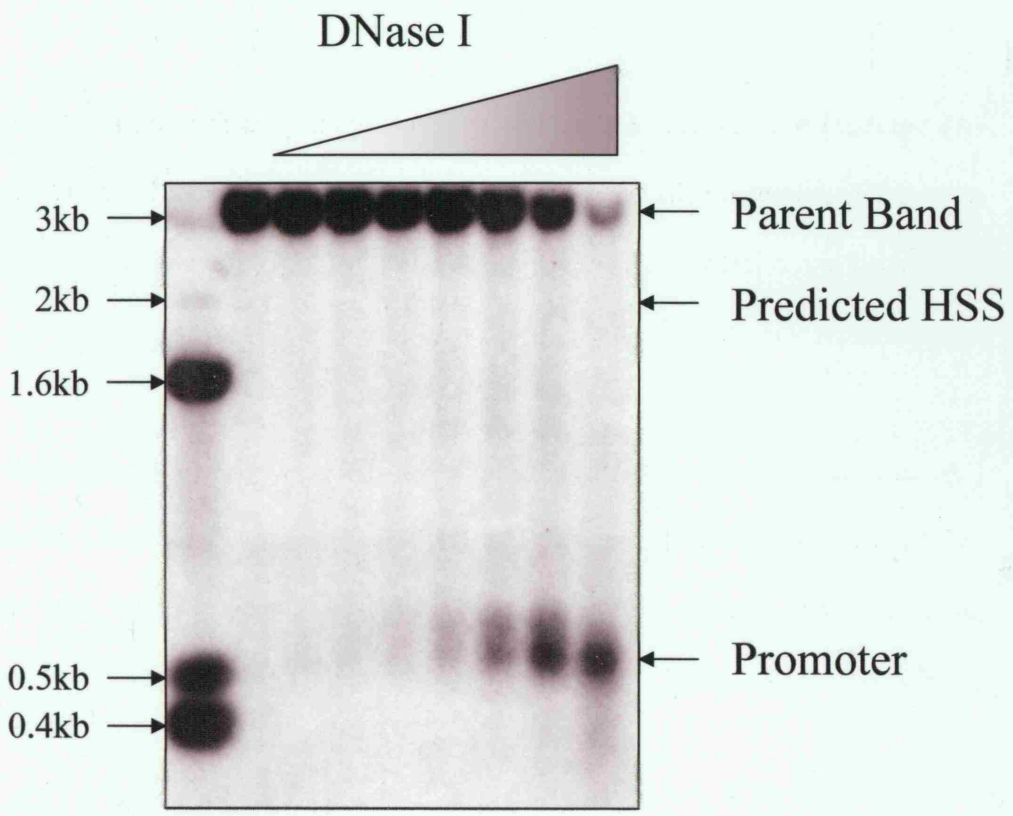
In this DNase I hypersensitivity assay, using a HinDIII digest and a probe from upstream of the predicted S/MAR site a 6.3kb parent band is generated (figure 28A). With increasing amounts of DNase I a second band of around 4.7kb is produced. This band corresponds to the DNase I HSS present within the promoter region of the hCD2 transgene (a diagram of the minigene is shown in figure 28B). No further DNase I HSSs

Figure 27. DNase I analysis of the hCD2 5' region in transgenic mice.

(A) Southern blot showing DNase I analysis of thymocytes from an MG4 transgenic mouse. The untreated cells are in lane 1, whereas cells treated with increasing amounts of DNase I are in lanes 2-8. The location of the 3kb parent band and promoter DNase I hypersensitive site are indicated.

(B) Schematic diagram of the hCD2 minigene construct which was used to generate the MG4 transgenic line. Hypersensitive sites (HSS) 1, 2 and 3 are indicated by arrows. The position of the BglII restriction sites and 5' cDNA probe are indicated. The expected position of the previously described 5' HSS is indicated. Open boxes represent exons.

A.


DNase I
3kb
2kb
1.6kb
0.5kb
0.4kb
Parent Band
Predicted HSS
Promoter

B.

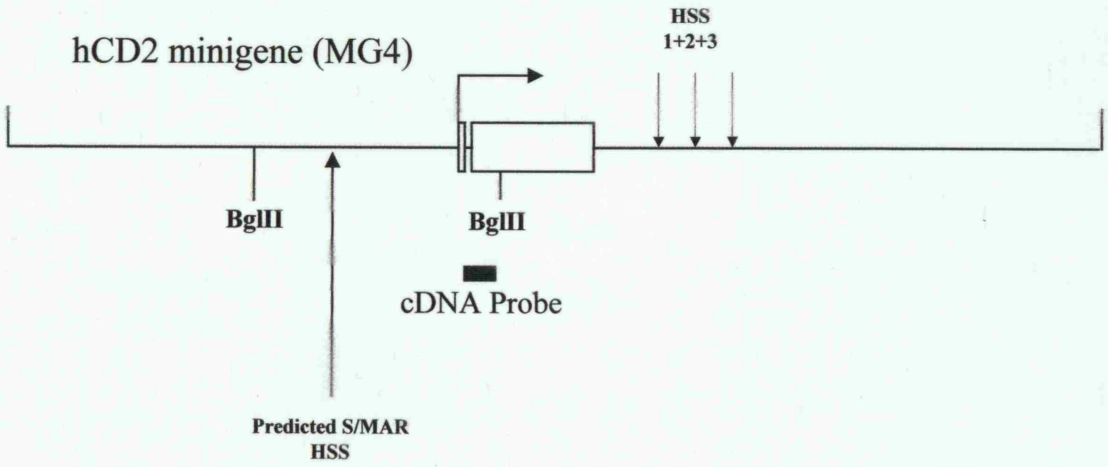
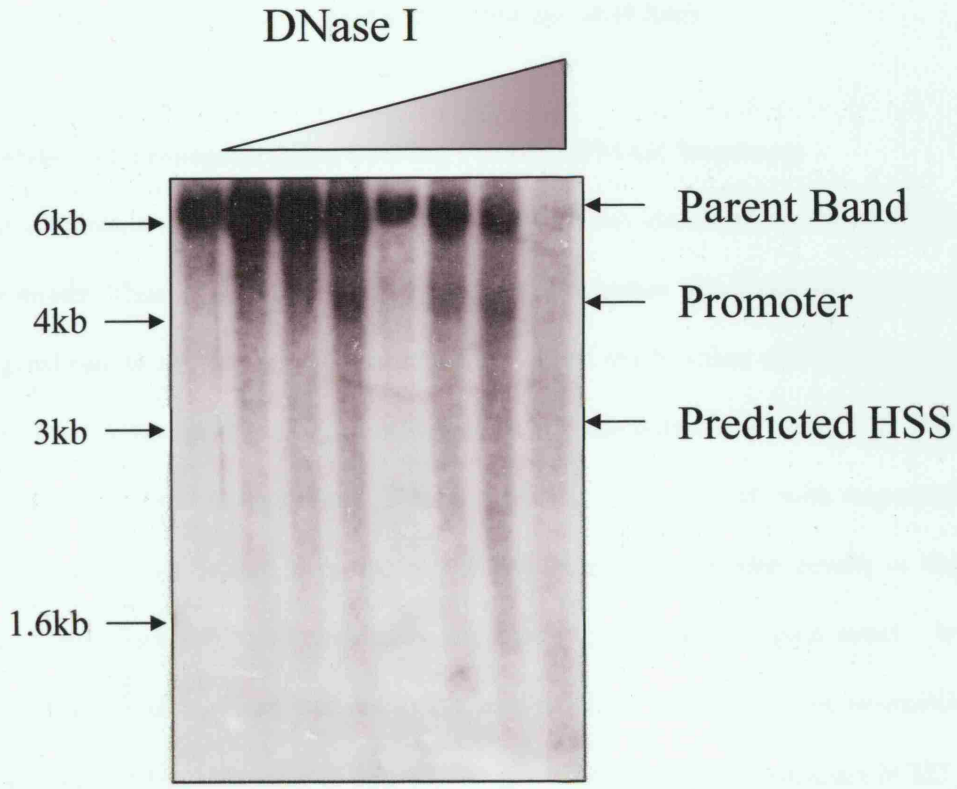

hCD2 minigene (MG4)
HSS
1+2+3
BglII
BglII
cDNA Probe
Predicted S/MAR
HSS

Figure 28. DNase I analysis of the hCD2 5' region in transgenic mice.

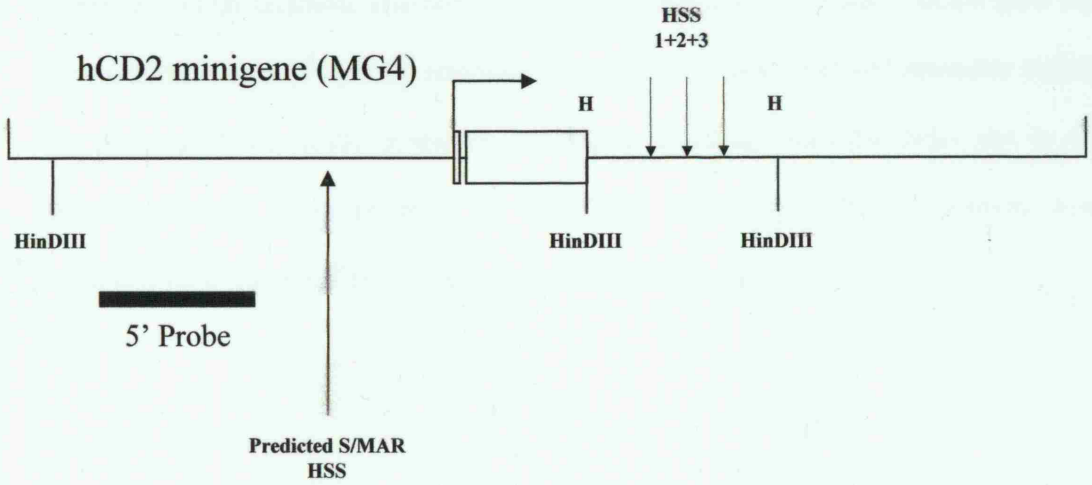
(A) Southern blot showing DNase I analysis of thymocytes from an MG4 transgenic mouse. The untreated cells are in lane 1, whereas cells treated with increasing amounts of DNase I are in lanes 2-8. The location of the 6kb parent band and promoter DNase I hypersensitive site are indicated.

(B) Schematic diagram of the hCD2 minigene construct which was used to generate the MG4 transgenic line. Hypersensitive sites (HSS) 1, 2 and 3 are indicated by arrows. The position of the HinDIII restriction sites and 5' probe are indicated. The expected position of the previously described 5' HSS is indicated. Open boxes represent exons.

A.


DNase I
6kb
4kb
3kb
1.6kb
Parent Band
Promoter
Predicted HSS

B.


HSS
1+2+3
hCD2 minigene (MG4)
H
H
HinDIII
HinDIII
HinDIII
5' Probe
Predicted S/MAR
HSS

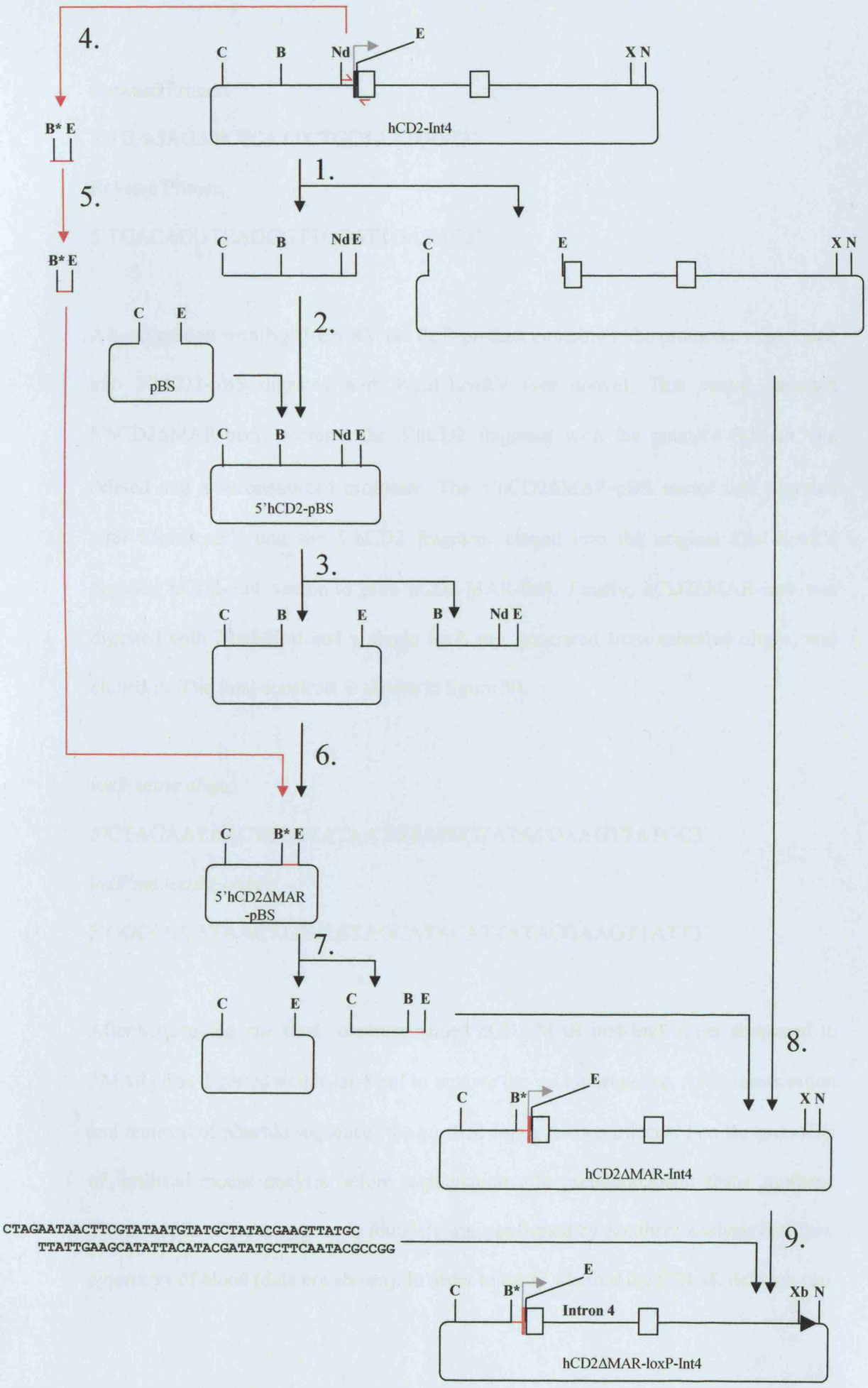
were detected. These results confirm that the upstream HSS within the putative S/MAR site does not form in transgenic mice (at least within the MG4 line).

6.3 Generation of Transgenic Mice Lacking Putative S/MAR Sequences

As describe above, a region with high S/MAR potential was identified upstream of the hCD2 promoter. Thus, it was interesting to determine whether this sequence serves a physiological role in regulating hCD2 expression. To test the function of this region in transgenic mice a construct was generated that lacked the putative S/MAR site. The sequences contained within the region -2596bp (BglII) to -522bp (NdeI) with respect to the initiation site were deleted from the hCD2-Int4 gene. This deletion results in the excision of the putative S/MAR sequences but leaves the promoter region intact. In addition, a single loxP site was inserted at the terminal 3' of the hCD2 gene to enable transgene copy number reduction if desired. To generate the ΔMAR construct hCD2-int4 was digested with ClaI-EcoRV and the 5' region of the gene was cloned into ClaI-EcoRV digested pBluescript (pBS) to give 5'hCD2-pBS (see figure 29 for a diagrammatic representation of the cloning). The only available sites for removing the putative S/MAR sequence also remove the promoter sequences. Thus, 5'hCD2-pBS was cut with BglII-EcoRV thereby removing the putative S/MAR site and promoter region. To reconstruct the hCD2 promoter, the region spanning from the NdeI site to the EcoRV site was amplified from the hCD2-Int4 vector by PCR. The primers were designed such that the NdeI site was mutated to a BglII site.

Figure 29. Description of the cloning to generate the ΔMAR construct.

The hCD2-Int4 was cut with ClaI/EcoRV (1) and cloned into pBluescript to generate the 5'hCD2-pBS vector (2). The potential S/MAR site and promoter sequences were deleted from 5'hCD2-pBS using a BglII/EcoRV digest (3). The promoter region was amplified by PCR (4) using primers that mutated the NdeI site to BglII (for primes sequences see materials and methods). The PCR product was cut with BglII/EcoRV (5) and cloned into BglII/EcoRV cut 5'hCD2-pBS to generate the 5'hCD2ΔMAR-pBS vector(6). The 5'hCD2ΔMAR-pBS vector was digested with ClaI/EcoRV to liberate the 5'hCD2 sequences (7) which were cloned into ClaI/EcoRV digested hCD2-Int4 to generate the hCD2ΔMAR-Int4 vector (8). The hCD2ΔMAR-Int4 vector was digested with XbaI/NotI (9) and a single loxP site (denoted as a black triangle) generated from annealed oligos was cloned into the XbaI/NotI hCD2ΔMAR-Int4 vector to generate the final vector hCD2ΔMAR-loxP-Int4. Open boxes represent exons. E=EcoRV. C=ClaI. B=BglII. Nd=NdeI. X=XbaI. N=NotI. B*=BglII site created by mutation of the original NdeI site.


4.
C
B
Nd
E
X N
hCD2-Int4
B* E
5.
1.
C
B
Nd E
C
E
X N
B* E
C
E
pBS
2.
C
B
Nd E
5'hCD2-pBS
3.
C
B
E
B
Nd E
6.
C
B* E
5'hCD2ΔMAR
-pBS
7.
C
E
C
B E
8.
C
B*
E
X N
hCD2ΔMAR-Int4
CTAGAATAACTTCGTATAATGTATGCTATACGAAGTTATGC
TTATTGAAGCATATTACATACGATATGCTTCAATACGCCGG
9.
C
B*
E
Intron 4
Xb N
hCD2ΔMAR-loxP-Int4

Forward Primer:

5'TGAGAGATCTCATTCTGCTATTGGC3'

Reverse Primer:

5'TCACAGGTCAGGGTTGTGTTGATAC3'

After digestion with BglII-EcoRV the PCR product containing the promoter was cloned into 5'hCD2-pBS digested with BglII-EcoRV (see above). This vector, denoted 5'hCD2ΔMAR-pBS, contains the 5'hCD2 fragment with the putative S/MAR site deleted and a reconstructed promoter. The 5'hCD2ΔMAR-pBS vector was digested with ClaI-EcoRV and the 5'hCD2 fragment cloned into the original ClaI-EcoRV digested hCD2-int4 vector to give hCD2-MAR-int4. Finally, hCD2ΔMAR-int4 was digested with XbaI-NotI and a single loxP site, generated from annealed oligos, was cloned in. The final construct is shown in figure 30.

loxP sense oligo:

5'CTAGAATAACTTCGTATAATGTATGCTATACGAAGTTATGC3'

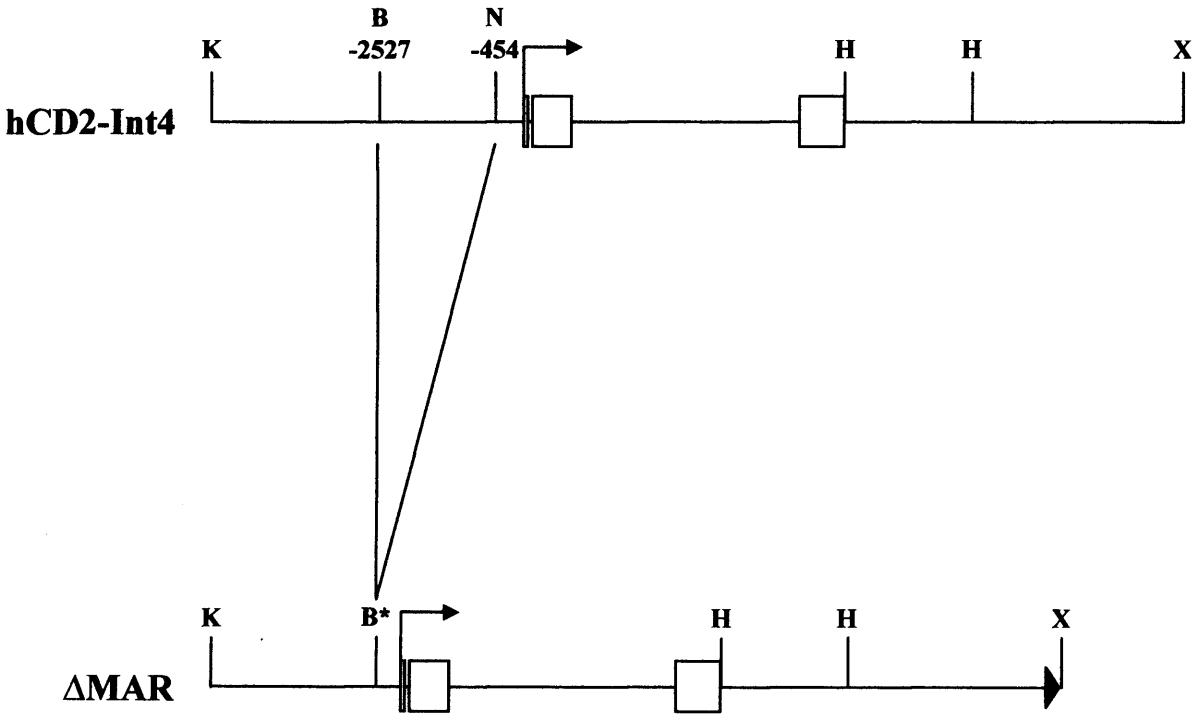
loxP anti-sense oligo:

5'GGCCGCATAACTTCGTATAGCATACATTATACGAAGTTATT3'

After sequencing, the final construct named hCD2-MAR-int4-loxP (later shortened to ΔMAR) was digested with ClaI-KpnI to remove the vector sequence. After linearisation and removal of plasmid sequences the purified construct was injected into the pronuclei of fertilized mouse oocytes before implantation into pseudopregnant foster mothers. The presence of the transgene in founders was confirmed by Southern analysis and flow cytometry of blood (data not shown). In order to study whether the S/MAR deletion can

Figure 30. The hCD2-Int4 and ΔMAR expression constructs.

The original hCD2-Int4 construct containing 5.5kb of downstream sequences is shown. The region containing the predicted MAR site lies between the BglII and NdeI sites (-2527 to -454 with respect to the transcriptional start site). The ΔMAR construct is also shown. A single loxP site was included at the terminal 3' of the construct (denoted as a black triangle). Open boxes represent exons. K=KpnI. B=BglII. N=NdeI. H=HindIII. X=XbaI. B*= BglII site created from mutation of the original NdeI site.


B
N
K
-2527
-454
H
H
X
hCD2-Int4
K
B*
H
H
X
ΔMAR

lead to variegation or other abnormalities in expression of the hCD2 transgene six independent founders were chosen and bred to CBA mice in order to establish six independent lines designated ΔMAR1-6. The line designated ΔMAR.4 did not transmit from the original founder.

Blood from each of the established ΔMAR lines was stained with anti-hCD4, anti-CD8 and anti-CD2 (figure 31). Analysis by flow cytometry revealed that all lines expressed high levels of hCD2. Moreover, 4 of the 5 lines displayed non-variegating expression of hCD2. The lines designated ΔMAR.2/5/6 were chosen for further analysis.

6.4 Transgene Copy Number in ΔMAR Transgenic Mice

Genomic tail DNA was isolated from ΔMAR transgenic mice and subjected to slot blot analysis to determine the approximate transgene copy number of each line. Slot blots were probed with the hCD2 cDNA and with the mCD2 cDNA to allow for loading normalisation. Figure 32 shows the calculated transgene copy number of each line. Transgenic lines of known hCD2 copy number were included as controls. Comparison to the lines of known copy number allows the approximate transgene copy number to be determined. The calculated copy numbers are ~70, ~125 and ~1-2 for the ΔMAR.2, ΔMAR.5 and ΔMAR.6 respectively.

6.5 Characterisation of hCD2 Expression on DP Thymocytes from ΔMAR Transgenic Mice

To characterise the transgene expression profile in ΔMAR transgenic mice, thymocytes were stained with anti-hCD4, anti-CD8 and anti-CD2. Analysis by flow cytometry revealed that all three ΔMAR lines express high levels of hCD2 on $CD4^+/CD8^+$ DP

Figure 31. Expression of hCD2 on $CD4^+$ peripheral blood T cells isolated from ΔMAR transgenic mice.

Blood from ΔMAR transgenic mice was stained with antibodies specific for CD4, CD8 and hCD2. Histograms show expression of hCD2 on CD4 single positive T cells.

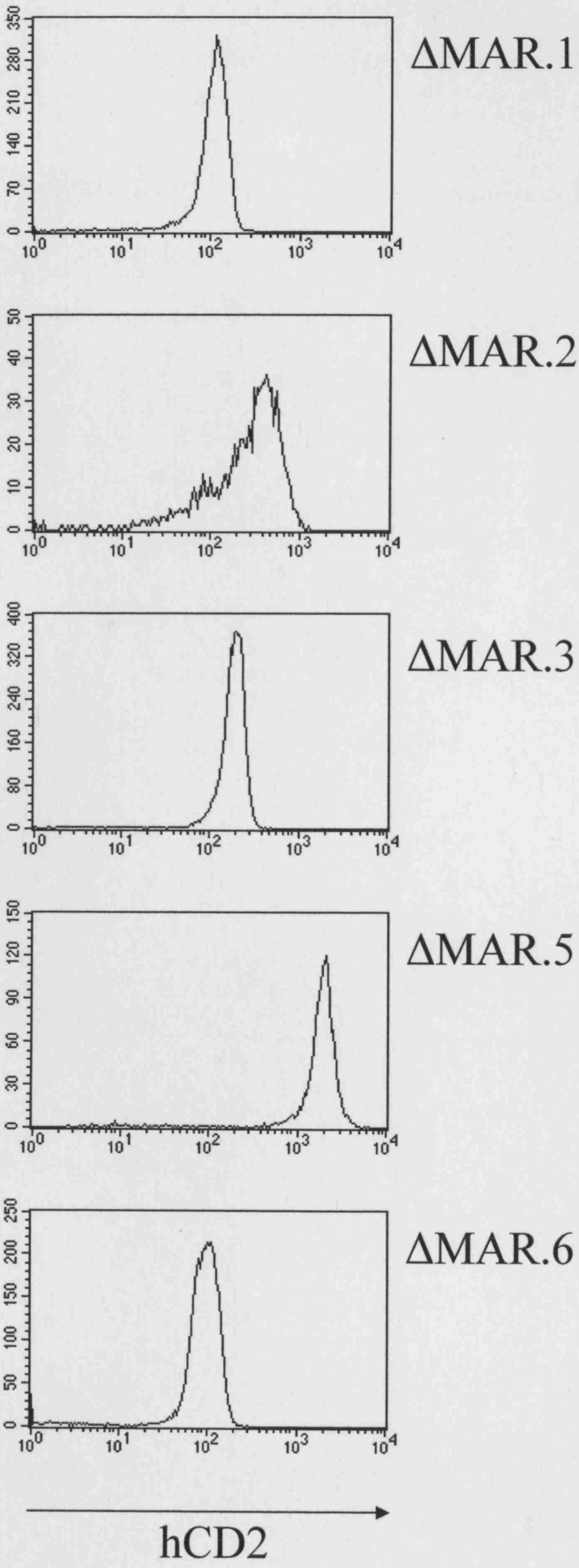
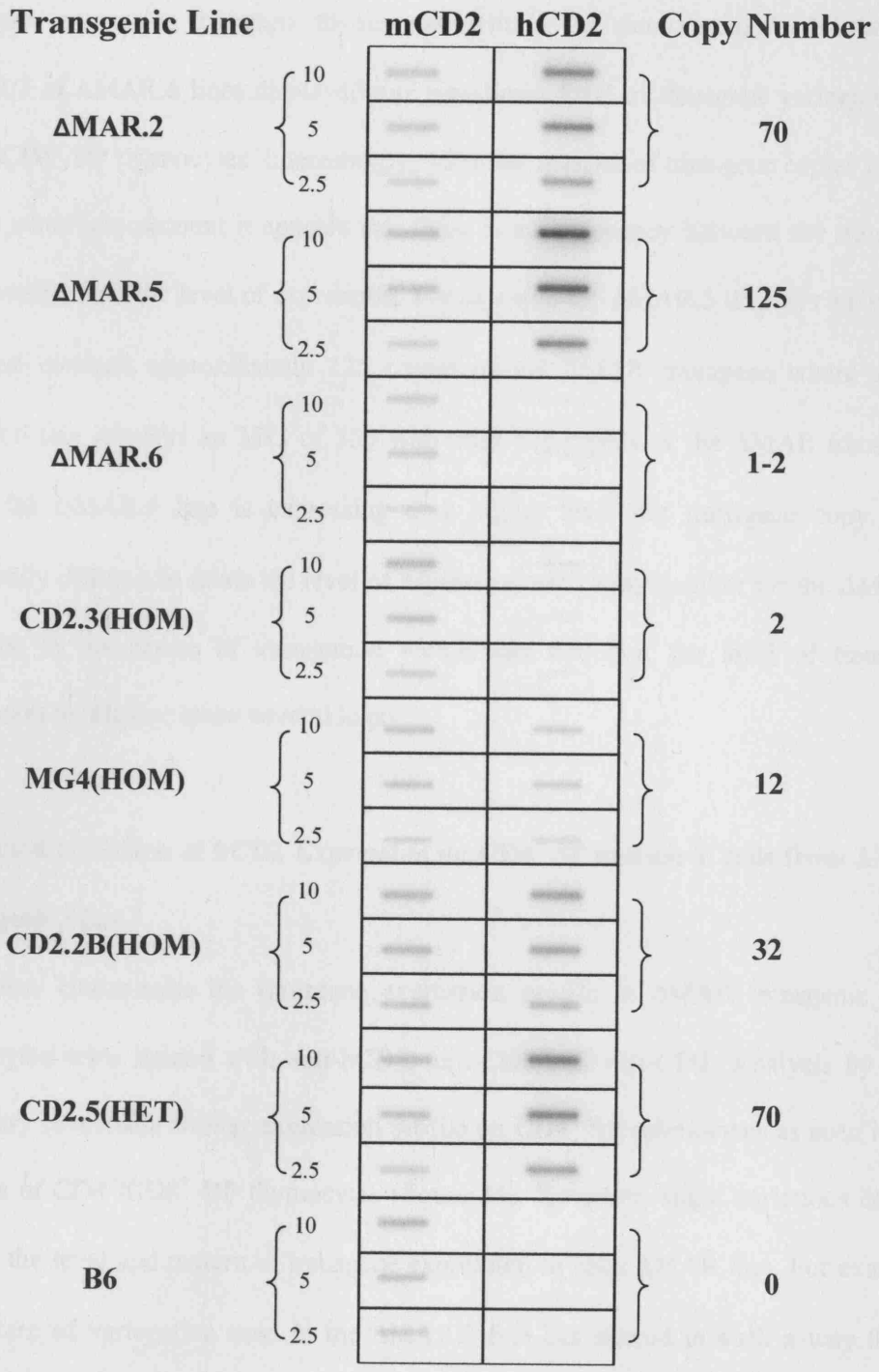

ΔMAR.1
ΔMAR.2
ΔMAR.3
ΔMAR.5
ΔMAR.6
hCD2

Figure 32. Slot blot analysis of transgene copy number in ΔMAR transgenic mice.

Samples containing 10μg, 5μg, and 2.5μg of genomic DNA were each equally loaded into two separate wells. Slots were probed with either hCD2 cDNA or mCD2 to allow normalisation of loading. Copy number was determined by comparison to the following transgenic lines CD2.3 (homozygous), MG4 (homozygous), CD2.2B (homozygous) and CD2.5 (heterozygous) which contain 2, 12, 32 and 70 copies of the hCD2 gene, respectively.


Transgenic Line
mCD2
hCD2
Copy Number
ΔMAR.2
10
5
2.5
70
ΔMAR.5
10
5
2.5
125
ΔMAR.6
10
5
2.5
1-2
CD2.3(HOM)
10
5
2.5
2
MG4(HOM)
10
5
2.5
12
CD2.2B(HOM)
10
5
2.5
32
CD2.5(HET)
10
5
2.5
70
B6
10
5
2.5
0

thymocytes (figure 33). Interestingly, as noted in the blood analysis, the ΔMAR.2 line displays a variegated phenotype (evident in this case a broad range in the levels of transgene expression between thymocytes within the same animal). Neither the ΔMAR.5 of ΔMAR.6 lines displayed any significant level of transgene variegation on $CD4^+/CD8^+$ DP thymocytes. Interestingly, when the number of transgene copies in each line is taken into account it appears that there is a discrepancy between the transgene copy number and the level of expression. For example the ΔMAR.5 displays an MFI of 753 and contains approximately 125 copies of the ΔMAR transgene where as the ΔMAR.6 line displays an MFI of 350 with only 1-2 copies of the ΔMAR transgene. Thus, the ΔMAR.6 line is expressing at a higher level per transgene copy. It is technically difficult to relate the level of expression with copy number for the ΔMAR.2 line due to the nature of variegation within this line (i.e. the level of transgene expression in this line spans several logs).

6.6 Characterisation of hCD2 Expression on $CD4^+$ SP splenic T cells from ΔMAR Transgenic Mice

To further characterise the transgene expression profile in ΔMAR transgenic mice splenocytes were stained with anti-hCD4, anti-CD8 and anti-CD2. Analysis by flow cytometry revealed a similar expression profile on $CD4^+$ SP splenocytes as seen in the analysis of $CD4^+/CD8^+$ DP thymocytes (figure 34). However, slight variations can be seen in the level and pattern of transgene expression in each ΔMAR line. For example the pattern of variegation seen in the ΔMAR.2 line has altered in such a way that a greater percentage of cells are expressing a high level of hCD2. Similarly the level of transgene expression in the ΔMAR.5 line has increased (as indicated by the MFI) on

Figure 33. Expression of hCD2 on double positive thymocytes isolated from ΔMAR transgenic mice.

Single cell suspensions from thymus were prepared from ΔMAR transgenic mice. Thymocytes were stained with antibodies specific for CD4, CD8 and hCD2. Dot plot analysis shows CD4 against CD8 staining and the circle defines the gated double positive population. Histograms show expression of hCD2 on double positive thymocytes. Values represent the mean fluorescence intensity of anti-hCD2-FITC.

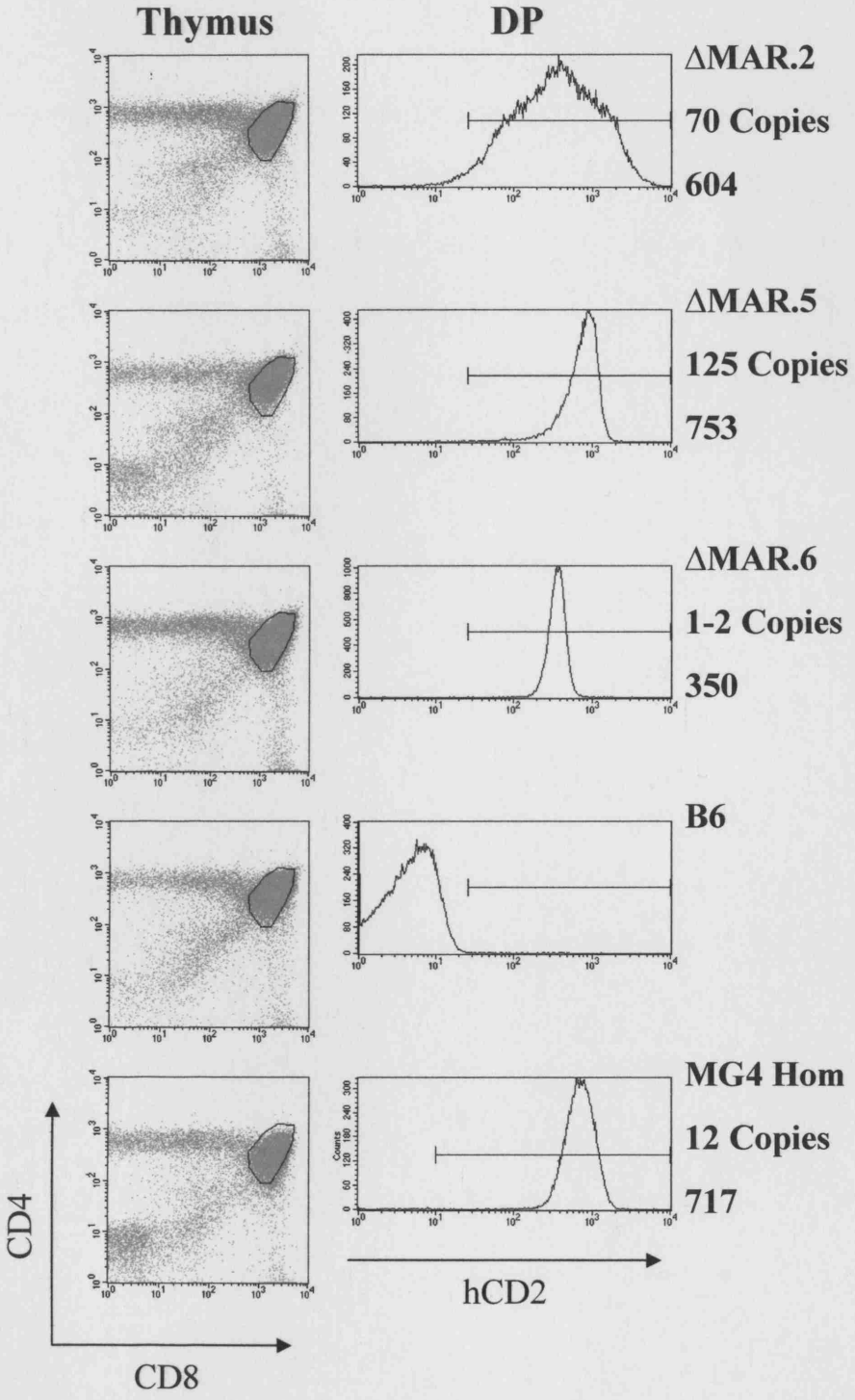
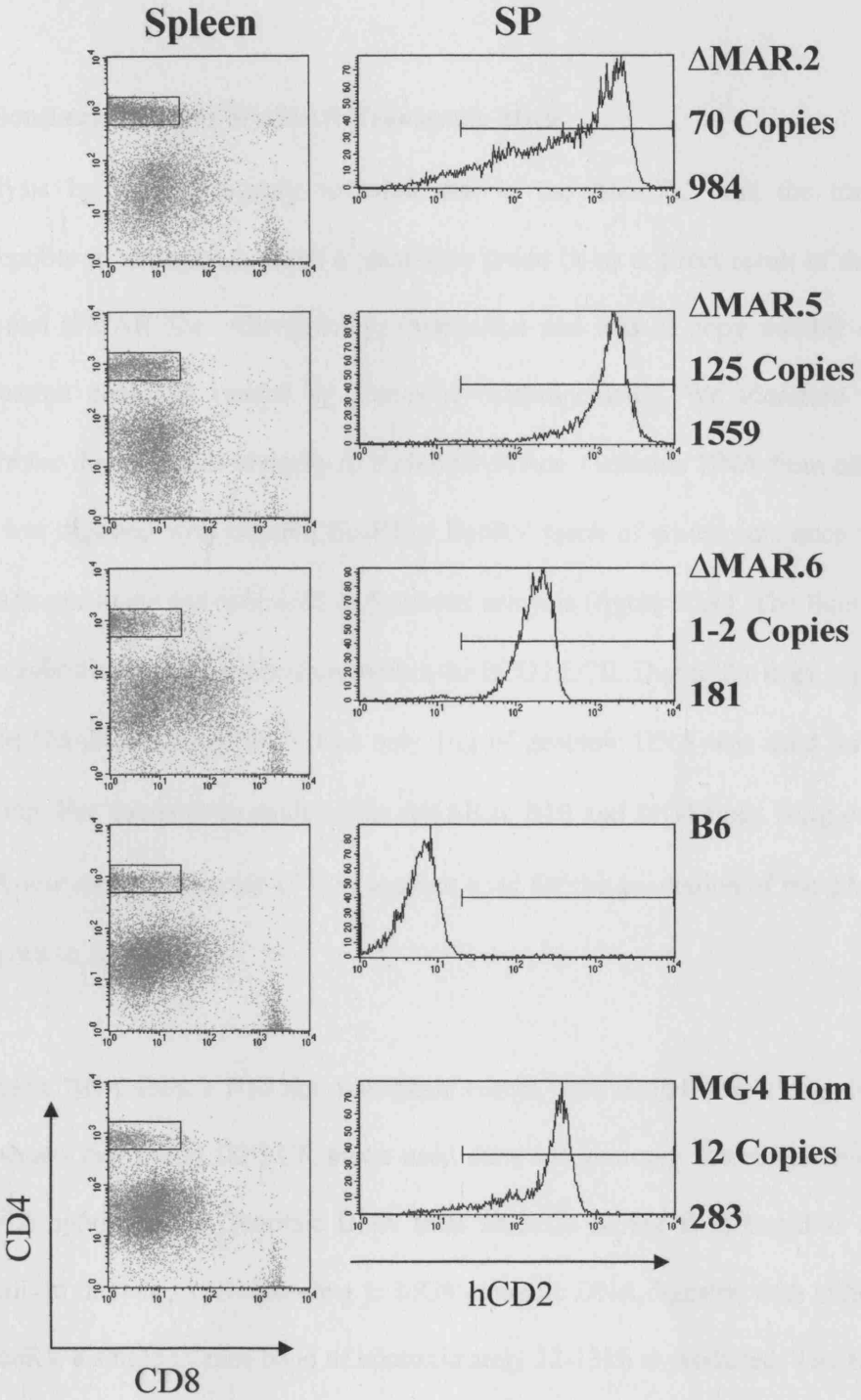

Thymus
DP
ΔMAR.2
70 Copies
604
ΔMAR.5
125 Copies
753
ΔMAR.6
1-2 Copies
350
B6
MG4 Hom
12 Copies
717
CD4
CD8
hCD2

Figure 34. Expression of hCD2 on single positive splenocytes isolated from ΔMAR transgenic mice.

Single cell suspensions from spleen were prepared from ΔMAR transgenic mice. Cells were stained with antibodies specific for CD4, CD8 and hCD2. Dot plot analysis shows CD4 against CD8 staining and boxes define gated populations. Histograms show expression of hCD2 on CD4 single positive splenocytes. Values represent the mean fluorescence intensity of anti-hCD2-FITC.


Spleen
SP
ΔMAR.2
70 Copies
984
ΔMAR.5
125 Copies
1559
ΔMAR.6
1-2 Copies
181
B6
MG4 Hom
12 Copies
283
CD4
CD8
hCD2

$CD4^+$ SP splenic T cells when compared to $CD4^+/CD8^+$ DP thymocytes. In contrast, the level of expression has decreased on the ΔMAR.6 line

6.7 Southern Analysis of ΔMAR Transgenic Mice

Analysis by flow cytometry revealed that in the ΔMAR.2 line the transgene is susceptible to variegation. Such a phenotype could be as a direct result of deleting the potential S/MAR site. Alternatively, variegation and loss of copy number-dependent expression could be caused by transgene rearrangements. We therefore sought to determine the transgene integrity in each ΔMAR line. Genomic DNA from each ΔMAR line was digested with BamHI, EcoRI or EcoRV (each of which cuts once within the ΔMAR construct) and subjected to Southern analysis (figure 35A). The Southern blots were hybridised with a probe from within the hCD2 LCR. Due to the high copy number in the ΔMAR.2 and ΔMAR.5 lines only 1μg of genomic DNA was used for Southern blotting. For Southern analysis of the ΔMAR.6, B10 and MG4 lines 10μg of genomic DNA was used. A diagram of the construct used for the generation of the ΔMAR mice is shown in figure 35B.

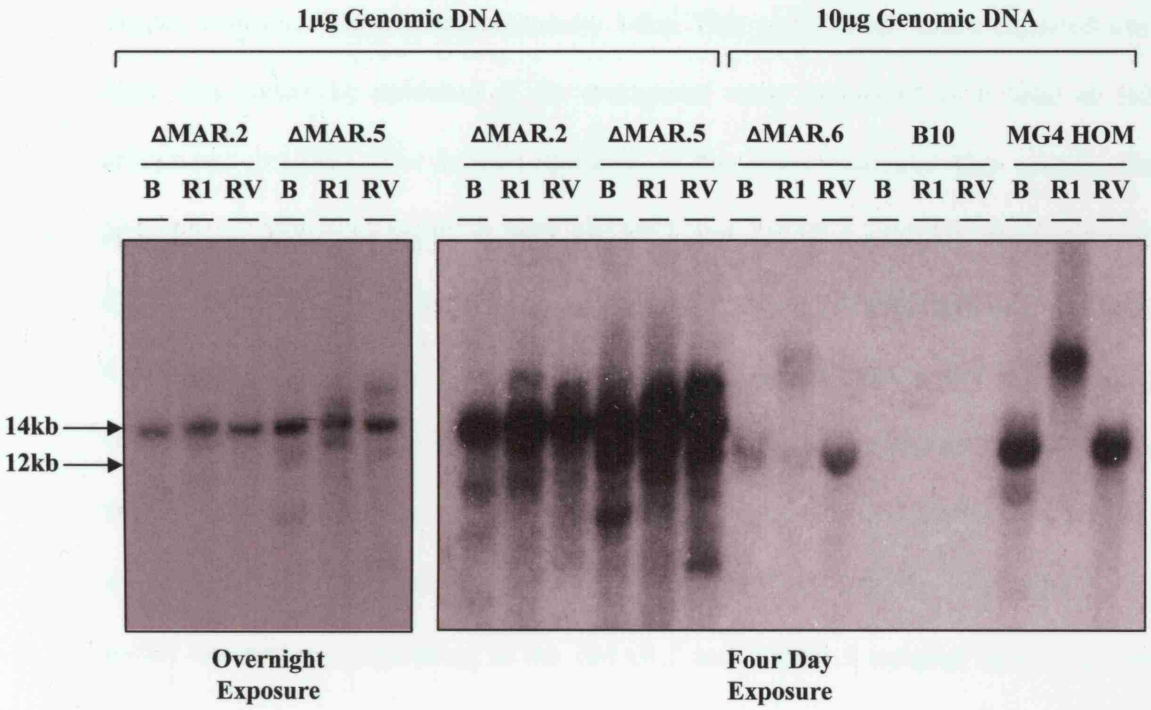
Genomic DNA from a B10 non-transgenic control was included as a negative control and shows that the hCD2 LCR probe used does not give any detectable level of non-specific hybridisation. Genomic DNA from an MG4 mouse is included as a positive control. In the lanes corresponding to MG4 genomic DNA digested with either BamHI or EcoRV a single intense band of approximately 12-13kb is produced. The size of this band corresponds to the expected size of the band that would be produced from multiple copies of the minigene integrated in a head to tail orientation (12.5kb). In the lane corresponding to MG4 genomic DNA digested with EcoRI a single band is detected. As

Figure 35. Southern analysis of ΔMAR transgenic mice.

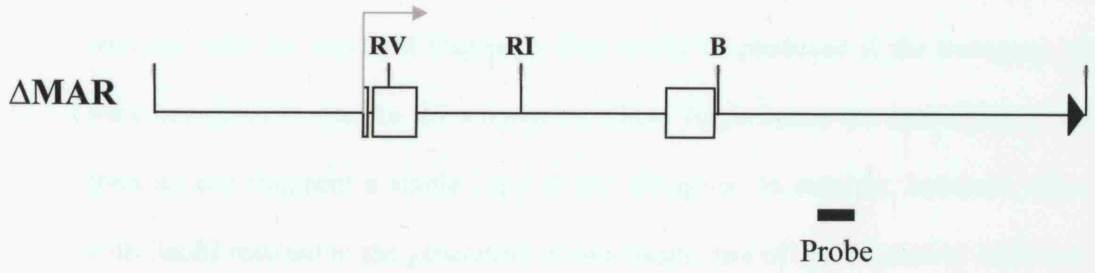
(A) Genomic DNA from ΔMAR transgenic mice was digested with BamHI, EcoRI or EcoRV. Genomic DNA from a B10 non-transgenic mouse included as a negative control. Genomic DNA from an MG4 transgenic mouse was included as a positive control. The probe is indicated below in (B). Left and right panels show short and long exposures of the same blot. The amount of DNA loaded in each lane is also indicated. B=BamHI. RI=EcoRI. RV=EcoRV.

(B) Schematic diagram of the ΔMAR construct which was used to generate the ΔMAR transgenic lines. Positions of relevant restriction sites and LCR probe are indicated. Open boxes represent exons. B=BamHI, RI=EcoRI, RV=EcoRV.

A.


1µg Genomic DNA
10µg Genomic DNA
ΔMAR.2
ΔMAR.5
ΔMAR.2
ΔMAR.5
ΔMAR.6
B10
MG4 HOM
B R1 RV B R1 RV B R1 RV B R1 RV B R1 RV B R1 RV B R1 RV
14kb
12kb
Overnight Exposure
Four Day Exposure

B.


ΔMAR
RV
RI
B
Probe

EcoRI does not cut within the minigene this fragment corresponds to the entire transgene array in this line. The left panel in figure 35A shows an overnight exposure of the same Southern blot. In both the ΔMAR.2 and ΔMAR.5 lines all three restriction digests produce a band of approximately 14kb. This corresponds to the expected size band that would be produced if the transgenes were integrated in a head to tail orientation (14.2kb). The longer exposure of the same Southern blot reveals the presence of additional bands in both ΔMAR.2 and ΔMAR.5 samples (right panel in figure 35A). The sizes of these bands do not correspond to the expected sizes of bands that would be produced if copies of the transgene were integrated in opposing orientation within the array (i.e. in tail to tail or head to head orientations). Furthermore, they do not correspond to end fragments; as only 1μg of DNA is loaded in each lane single copy end fragments would not be detectable. Therefore, the additional bands within the lanes corresponding to the ΔMAR.2 and ΔMAR.5 samples most probably originate from multiple rearranged transgene copies from within the array. The presence of rearranged transgene copies in the ΔMAR.2 and ΔMAR.5 transgenic lines may influence expression of hCD2. Thus, caution must be used when interpreting the data obtained from these two transgenic lines. Digestion of genomic DNA from the ΔMAR.6 transgenic line with BamHI and EcoRV generated in both cases a single band of approximately 10kb. The size of this band is too small to have originated from multiple transgene copies integrated in a head to tail orientation. Furthermore the sizes do not correlate with the expected fragments that would be produced if the transgene copies were integrated in a tail to tail orientation. These single bands are most likely produced from an end fragment a single copy of the transgene. In contrast, however, digestion with EcoRI resulted in the generation of two bands, one of approximately 10kb and one greater than 20kb. The presence of two bands in this digest suggests that there is more

than one copy of the transgene integrated in this line, although possibly one being an incomplete copy. This observation is in agreement with the copy number analysis, which suggested that the line contains 1-2 copies of the ΔMAR transgene.

6.8 Discussion

Using a computational approach we have sought to identify additional regulatory elements within the hCD2 locus. The S/MAR prediction program MAR-Wiz identified a sequence with a high S/MAR site potential within the region -2500 to -500 bp, with respect to the hCD2 initiation site. It is important to remember that such programs can only make predictions based on sequence composition and do not prove physiological function. However, the MAR-Wiz program has been suggested to under-predict the presence of functional S/MAR sites. Furthermore, although the predicted hCD2 S/MAR site may not necessarily function as a S/MAR element *in vivo*, the S/MAR finder program does reveal that its sequence composition is unquestionably different from the rest of the hCD2 locus. Thus we sought to examine whether this sequence was important for the regulation of hCD2 expression.

6.8.1 DNase I Analysis of the Predicted S/MAR Sequence

Previously, in human cell lines, a DNase I HSS has been identified upstream of the hCD2 promoter, within the region reported here to contain a potential S/MAR site [491]. As DNase I HSS's are thought to represent binding sites of regulatory proteins this strengthens the possibility that the region within the predicted S/MAR element contains regulatory sequences.

An initial DNase I hypersensitivity assay on thymocytes from a hCD2 transgenic mouse failed to reveal the presence of the 5'HSS in our system. Importantly, however, a HSS was shown to form at the hCD2 promoter, confirming that the DNase I hypersensitivity assay had worked from a technical point of view. One potential caveat of this DNase I

hypersensitivity assay is that the promoter lies between the probe and the expected position of the DNase I HSS within the predicted S/MAR site. Thus, the strong promoter DNase I HSS may mask the presence of the upstream HSS if it is considerably weaker. Perhaps relevant to this notion Wotton et al. detected the 5' HSS using a different digest and probe than reported here [491]. The DNase I HSS analysis was therefore repeated utilising a different restriction digest and probe. Unfortunately, the sequence upstream of the promoter is relatively repetitive and as a result the probe used gives a high background. However, this analysis also failed to reveal the presence of a HSS within the putative S/MAR site. Thus it appears that the 5' HSS does not form in MG4 transgenic mice. It is worth noting that the original analysis was preformed using human cell lines, which raises the possibility that the 5' HSS may not form when the hCD2 gene is integrated into the murine genome as a transgene.

6.8.2 Analysis of hCD2 Transgenic Mice Lacking the Potential S/MAR Element

To determine whether the predicted S/MAR site serves any function in regulating hCD2 expression, transgenic mice were generated utilising a construct in which the predicted S/MAR site had been deleted.

Analysis of hCD2 expression on peripheral T cells from blood revealed that all founder mice containing the ΔMAR transgene expressed hCD2, indicating that the S/MAR site was not absolutely necessary for hCD2 expression.

Subsequent flow cytometry of the ΔMAR.2, ΔMAR.5 and ΔMAR.6 lines revealed a discrepancy between transgene copy-number and the level of expression (i.e. the ΔMAR.5 line expressed at a lower level per transgene copy than the ΔMAR.6 line).

Furthermore, the ΔMAR.2 line exhibited a variegated expression profile. However, Southern analysis revealed that the ΔMAR.2 and ΔMAR.5 transgenic lines contained rearranged copies of the ΔMAR transgene. Transgene rearrangements are known to have deleterious effects on transgene expression. Furthermore, transgene rearrangements can even induce variegation of hCD2 constructs linked to the full hCD2 LCR in transgenic mice (D. Kioussis unpublished observations). However, the number of rearranged transgene copies in each line is low. As it is not known whether the presence of these rearranged transgene copies within the array has any influence on the pattern of hCD2 expression in these lines, caution must be taken when drawing conclusions from ΔMAR.2 and ΔMAR.5 transgenic mice regarding the affect of deleting the potential S/MAR site from the hCD2 gene. However, the fact that variegation was not seen in 4 of the 5 transgenic lines indicates that the S/MAR site does not play an essential role in protecting against position effects. Two additional ΔMAR transgenic lines have recently been generated, neither of which display variegated hCD2 expression (data not shown). This provides additional support for the conclusion that the predicted S/MAR site is not essential to protect against PEV in transgenic mice. It is formally possible that the transgenes in these lines have not integrated into regions within the genome that would induce PEV. However, this seems unlikely as in general variegating lines are more frequently obtained than those that do not variegate (for example in sections 3.3-3.5 7 independent lines were generated only one of which did not display variegation). Thus it appears that the potential S/MAR site serves no role in protecting hCD2 transgenes from PEV in transgenic mice.

6.8.3 What is the Role of the Putative S/MAR Site?

Previously, the 2kb LCR has been shown to be sufficient to direct position-independent, copy number-dependant transgene expression upon transgenes under the control of heterologous promoters. Furthermore, the hCD2 LCR is able to confer copy-number dependant position-independent expression upon linked genes when included in lentiviral vectors [492, 493]. Therefore, it appears that the 'LCR defining' attributes reside solely within this 2kb sequence, and are independent of the promoter used (provided the promoter is active in T cells). Thus, if the potential S/MAR site represents a regulatory element important for hCD2 expression what could its role be if the 2kb LCR is sufficient for correct expression? There are a number of possibilities. For example, the S/MAR site could represent an enhancer element. Comparing ΔMAR transgenic mice to hCD2-Int4 transgenic mice (which only differ by the presence of the potential S/MAR site) may reveal a difference in the expression level. However, due to the potential transgene rearrangements within the ΔMAR.2 and ΔMAR.5 transgenic lines caution must be used when interpreting data obtained from analysis of these mice.

Alternatively, the predicted S/MAR element may not be important in the regulation of hCD2 expression, but rather functions as a barrier element to prevent the powerful LCR from influencing expression of proximal genes upstream of the endogenous hCD2 locus. With this possibility in mind it is interesting to note that knocking the hCD2 LCR into the CD8 locus results in the ectopic expression of both the CD8α and CD8β genes in $CD4^+$ T cells. This provides evidence that the hCD2 LCR can influence genes proximal to the site of integration.

6.8.4 Summary

In conclusion, using a computational approach we have identified a potential S/MAR element within the hCD2 locus. However, the region spanning the putative S/MAR site failed to form a DNase I HSS in transgenic mice. In addition, deletional analysis failed to reveal any essential role for this region in the regulation of hCD2 expression in transgenic mice.

Chapter Seven

Identification of Non-Coding Transcripts within the hCD2 LCR

7.1 The Role of Non-Coding Transcription

Intergenic or non-coding transcription was regarded until recently as non-functional. It has been proposed that non-coding transcription is initiated by chance from 'pseudo promoters' lying within domains of open chromatin. However, it has become increasingly apparent that non-coding transcription does play an important role in the regulation of a number of gene complexes. For example, expression from the *Drosophila* bithorax complex is regulated by intergenic transcription, which is thought to function by defining domains of cis-regulatory activity in early embryogenesis [494]. In addition, non-coding transcription within the β-globin LCR and also between the globin genes has been shown to have a direct role in the regulation of the β-globin gene complex in mice [389-391]. This intergenic transcription is important for the establishment of an open chromatin structure that is conducive to globin gene expression and is essential for directing the developmental timing of globin gene switching [390]. How this regulation is achieved is not known in detail, however, it is thought that the act of transcription itself may result in the modification of the chromatin structure, priming the locus for globin transcription. Indeed, the passage of RNA Pol II through a chromatin template has been shown to alter its composition, primarily via the displacement of histone H1 and heterodimers of histones H2A/H2B [157, 158]. In addition, histone acetyltransferases (HATs) are able to bind to the C-terminal domain of RNA Pol II during elongation. It is therefore possible that RNA Pol II-associated HATs are able to acetylate the histone tails as the polymerase progresses along the template, thus coupling transcription to chromatin remodelling [157, 495].

Intergenic transcription has been also identified within the human and murine *IL-4/IL-13* gene cluster, within the MHC class II locus control region and the murine *Igh* locus [496-499].

Thus, to investigate the possibility that non-coding transcription contributes to the mechanism by which the hCD2 LCR generates an open chromatin domain in T cells, the presence of hCD2 LCR transcripts in hCD2 expressing cells was determined.

7.2 RNase Protection analysis of Non-Coding Transcripts within the hCD2 LCR

To generate templates for the RNase protection assay the following primers were used to amplify a 183 bp fragment from the hCD2 LCR.

Forward Primer:

5'GACTGGATCCGATCACCTGAGGTC3'

Reverse Primer:

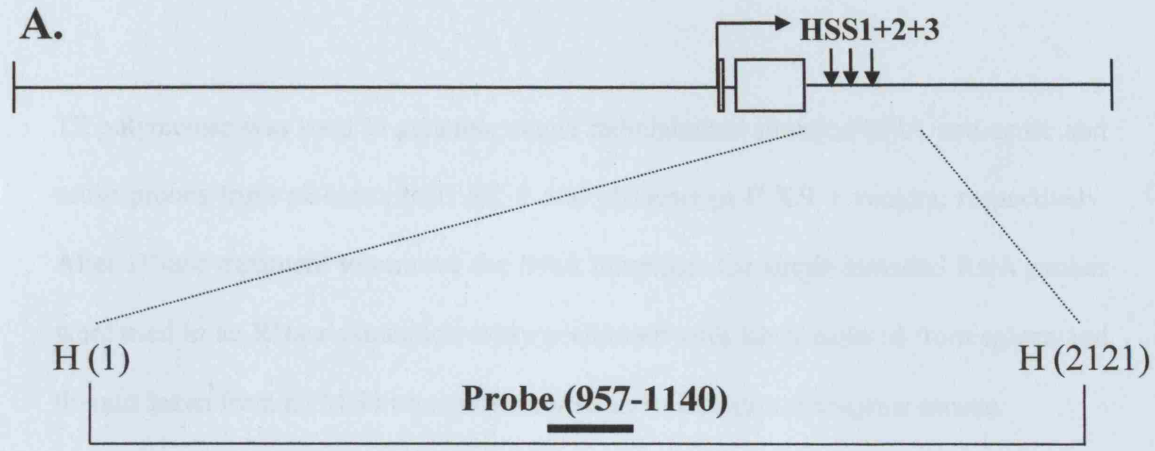
5'GATCAAGCTTCTCACCTCACTGC3'

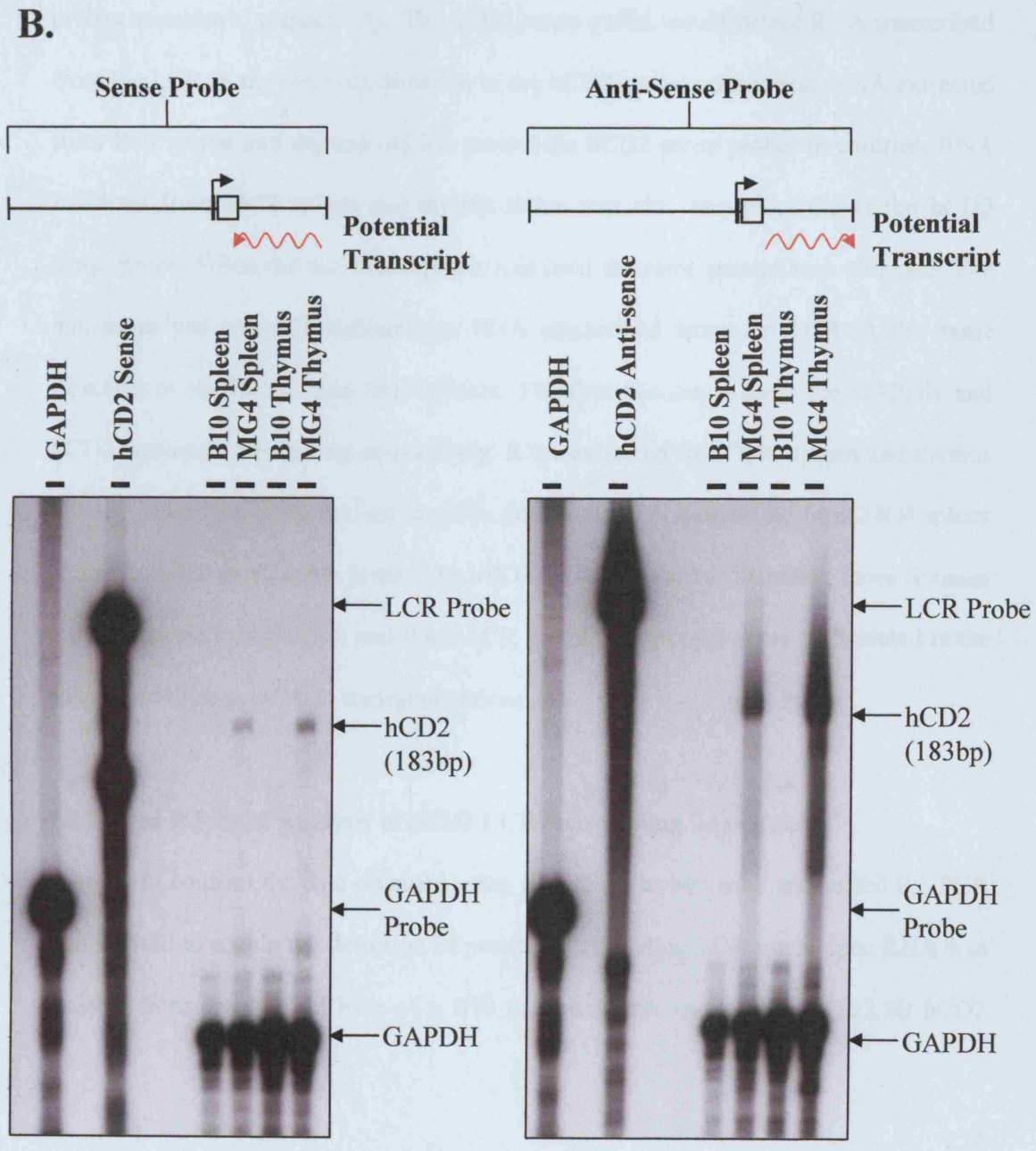
The resulting PCR product was digested with BamH1/HindIII and cloned into BamH1/HindIII cut pBluescript II SK + and pBluescript II KS + to generate vectors designated hCD2 LCR pBluescript II SK + and hCD2 LCR pBluescript II KS +, respectively. The 183bp LCR fragment is localised within the region 957-1140bp downstream of the 5' HinDIII site of the 2.1kb LCR (figure 36A).

Figure 36. RNase protection analysis to identify the presence of LCR transcripts in hCD2 transgenic mice.

(A) The hCD2 minigene is shown, and below an expanded view of the 2.1kb LCR, as demarcated by two HindIII sites (denoted H). The approximate location of the probe is represented by a black bar and the distance from the upstream HindIII site, that demarcates the start of the LCR, is indicated.

(B) RNase protection was performed on RNA isolated from an MG4 transgenic mouse and a non-transgenic mouse. A schematic diagram is also shown to represent the direction of the LCR transcripts (indicated by a red wavy arrow). Positions of free and protected probes are also indicated.


A.
HSS1+2+3
H (1)
H (2121)
Probe (957-1140)


B.
Sense Probe
Potential Transcript
Anti-Sense Probe
Potential Transcript
GAPDH
hCD2 Sense
B10 Spleen
MG4 Spleen
B10 Thymus
MG4 Thymus
GAPDH
hCD2 Anti-sense
B10 Spleen
MG4 Spleen
B10 Thymus
MG4 Thymus
LCR Probe
hCD2 (183bp)
GAPDH Probe
GAPDH
LCR Probe
hCD2 (183bp)
GAPDH Probe
GAPDH

T7 polymerase was used to generate single radiolabelled stranded RNA anti-sense and sense probes from pBluescript II SK + and pBluescript II KS + vectors, respectively. After DNase treatment to remove the DNA templates the single stranded RNA probes were used in an RNase protection assay performed with RNA isolated from spleen and thymus taken from an MG4 transgenic mouse and a B10 non-transgenic mouse.

Figure 36B shows RNase protection analysis performed on RNA isolated from a MG4 transgenic mouse and a B10 non-transgenic mouse. The GAPDH and hCD2 sense free probes are shown, respectively. The hCD2 sense probe would detect RNA transcribed from the LCR in the opposite direction to the hCD2 gene transcription. RNA extracted from B10 spleen and thymus did not protect the hCD2 sense probe. In contrast, RNA extracted from MG4 spleen and thymus tissue was able to protect the 183bp hCD2 sense probe. When the anti-sense probe was used the same pattern was obtained. The anti-sense probe would hybridise to RNA transcribed from the LCR in the same direction as the hCD2 gene transcription. The first two lanes show the GAPDH and hCD2 anti-sense free probes respectively. RNA extracted from B10 spleen and thymus did not protect the hCD2 anti-sense probe. In contrast, RNA extracted from MG4 spleen and thymus tissue was able protect the hCD2 anti-sense probe. Together, these findings indicate that both anti-sense and sense LCR specific transcripts could be detected in the spleen and thymus of MG4 transgenic mice.

7.3 Nested RT-PCR Analysis of hCD2 LCR Non-Coding Transcripts

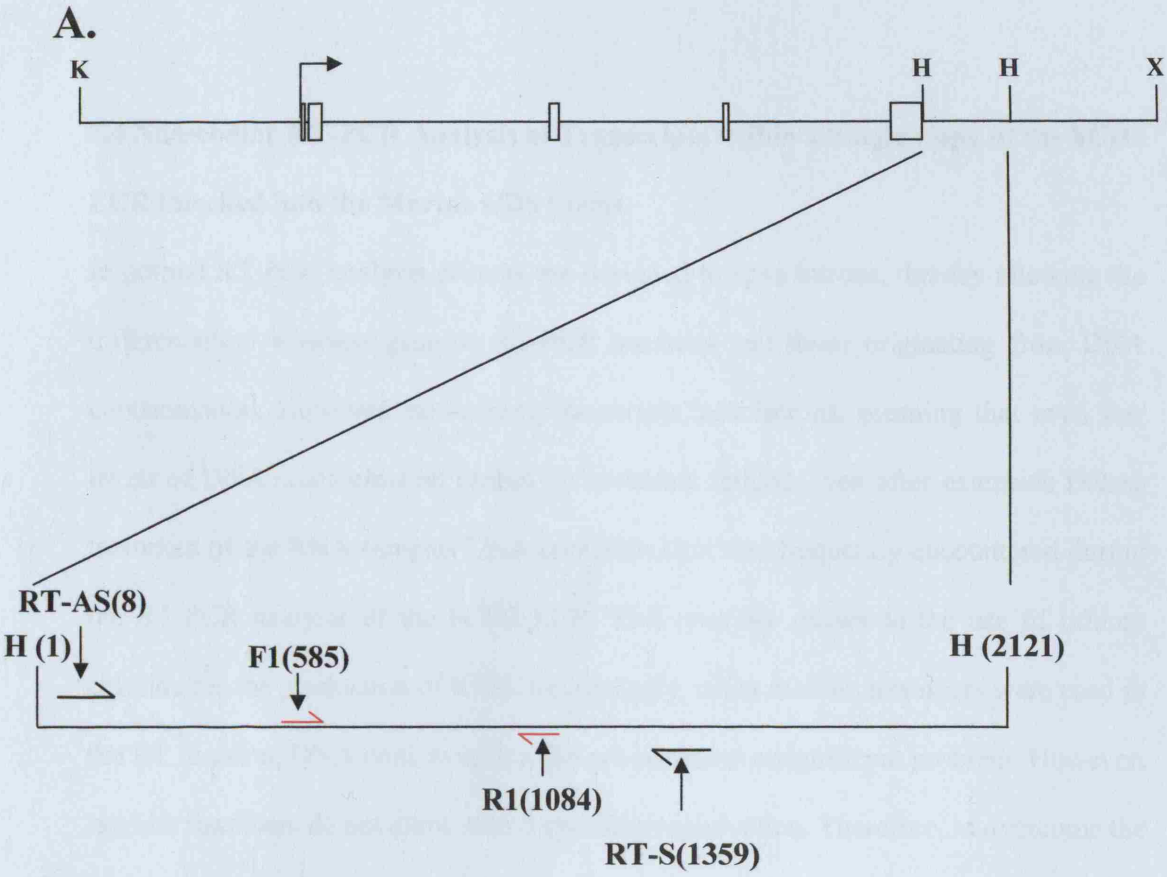
In order to confirm the data obtained using RNase protection analysis, nested RT-PCR was utilised to enable the detection of potential non-coding LCR transcripts. RNA was isolated from thymus and liver of a B10 non-transgenic mouse and a CD2.2B hCD2

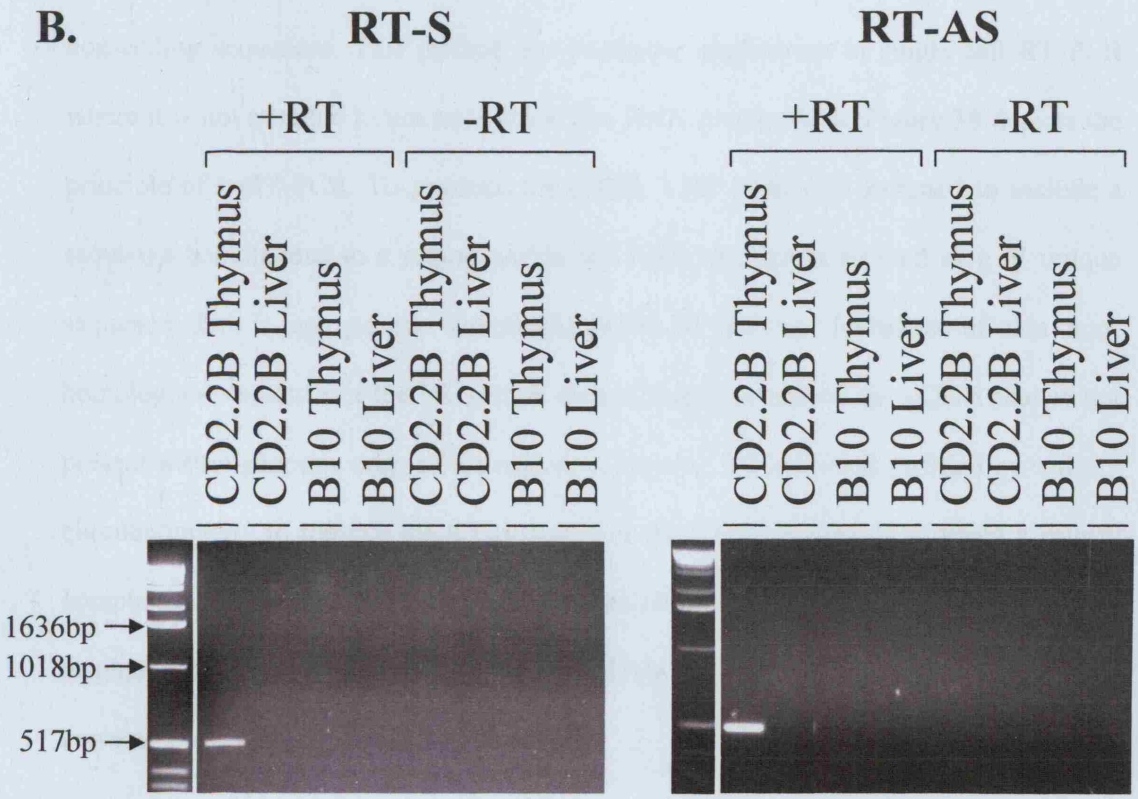
transgenic mouse. The heterozygous hCD2.2B transgenic line contains 16 copies of the hCD2 maxigene (figure 1A). Isolated RNA was treated with DNase to remove any residual DNA contamination. To generate cDNAs from the isolated RNA two different primers were used. The first primer (RT-AS see figure 37A) would hybridise to RNA that has been transcribed in the reverse orientation to the hCD2 gene transcription. The second primer (RT-S see figure 37A) hybridises to RNA that has been transcribed from the LCR in the same direction as the hCD2 gene and could also detect any read-through from the coding region of hCD2 gene itself. Both reactions were performed in the presence or absence of reverse transcriptase to exclude any DNA contamination. The cDNA was then used in a PCR reaction using LCR specific primers (denoted as F1 and R1 see figure 37A). When the RT-AS-cDNA from the CD2.2B thymus was used as a PCR template using LCR specific primers an amplification product of 500bp was generated, indicating that transcription through this region of the LCR had occurred in the opposite orientation to the hCD2 gene (figure 37B). When an identical PCR reaction was performed using the RT-S-cDNA as a template, generated from hCD2.2B thymus RNA, a specific PCR product of 500bp was produced, indicating that transcription through this region of the LCR had occurred in the sense direction (figure 37B). Importantly, no PCR products were generated when the RT negative controls were used as PCR templates, ruling out the possibility of DNA contamination. In addition, when the hCD2.2B liver cDNAs or B10 thymus and liver cDNAs were used as templates no PCR products were generated.

Figure 37. Nested RT-PCR analysis to identify LCR transcripts in hCD2 transgenic mice.

(A) The hCD2 maxigene is shown, and below an expanded view of the 2.1kb LCR, as demarcated by two HindIII sites. The approximate positions of the primers used are indicated (for primer sequences see materials and methods). Numbers in brackets represent the position of the primers in relation to the upstream HindIII demarcating the start of the LCR. RT-AS and RT-S refer to the primers used for cDNA generation. F1 and R1 refer to the primers used in the subsequent PCR reaction. H=HinDIII. K=KpnI. X=Xba1.

(B) Nested RT-PCR analysis was performed on RNA isolated from a hCD2.2B transgenic mouse and a B10 non-transgenic mouse. +RT and –RT refers to reactions performed in the presence or absence of reverse transcriptase, respectively.


A.
K
H
H
X
RT-AS(8)
H (1)
F1(585)
H (2121)
R1(1084)
RT-S(1359)

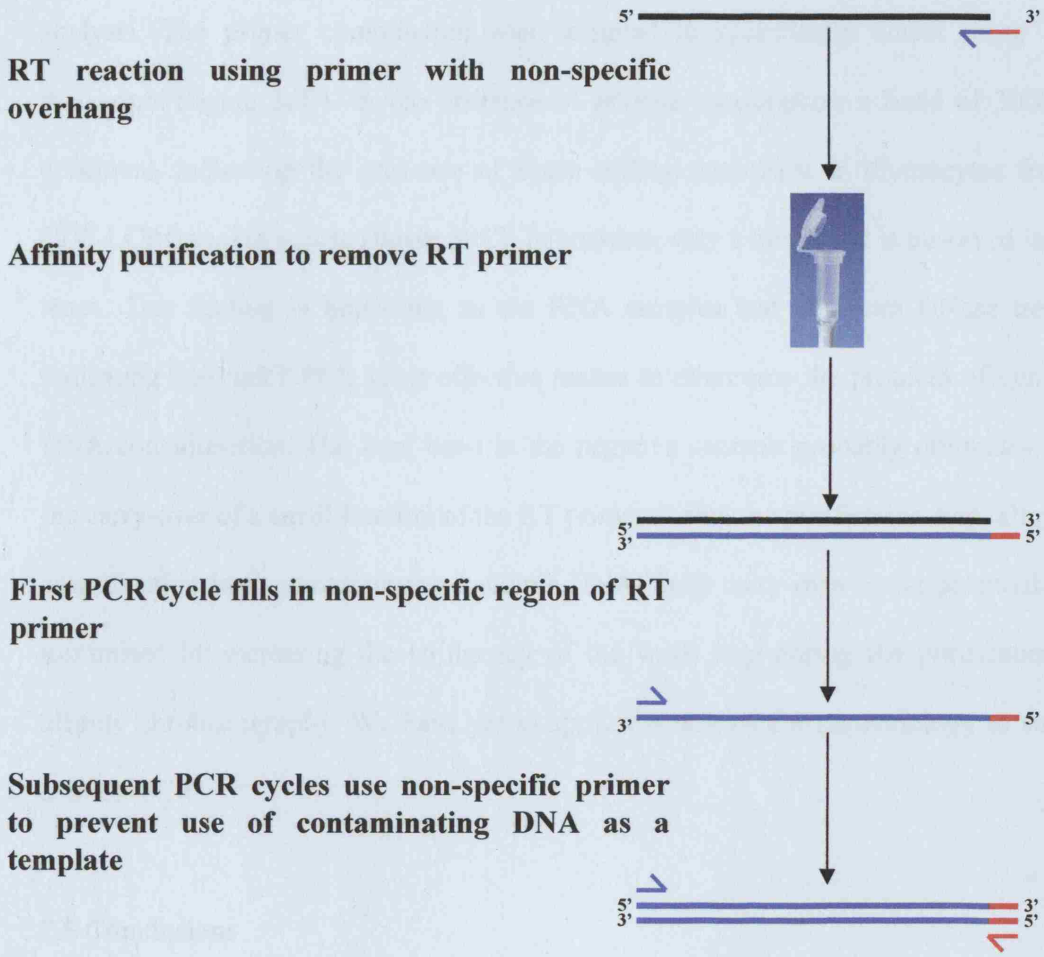

B.
RT-S
RT-AS
+RT
-RT
+RT
-RT
CD2.2B Thymus
CD2.2B Liver
B10 Thymus
B10 Liver
CD2.2B Thymus
CD2.2B Liver
B10 Thymus
B10 Liver
CD2.2B Thymus
CD2.2B Liver
B10 Thymus
B10 Liver
CD2.2B Thymus
CD2.2B Liver
B10 Thymus
B10 Liver
1636bp
1018bp
517bp

7.4 Non-coding RT-PCR Analysis of Transcripts within a Single Copy of the hCD2 LCR knocked into the Murine CD8 Locus.

In normal RT-PCR analysis primers are designed to span introns, thereby allowing the differentiation between genuine RT-PCR products and those originating from DNA contamination. However, non-coding transcripts lack introns, meaning that even low levels of DNA contamination cannot be excluded. Indeed, even after extensive DNase treatment of the RNA samples DNA contamination was frequently encountered during the RT-PCR analysis of the hCD2 LCR. This possibly relates to the use of lithium chloride for the production of RNA. Interestingly, when random hexamers were used in the RT reaction, DNA contamination did not represent a significant problem. However, random hexamers do not allow strand specific amplification. Therefore, to overcome the problem of DNA contamination a method, termed non-coding RT-PCR (ncRT-PCR), was developed that enables the specific detection of RNA transcripts originating from non-coding sequences. This method has particular application in single-cell RT-PCR where it is not possible to generate DNA free RNA preparations. Figure 38 depicts the principle of ncRT-PCR. To generate the cDNA a RT primer is designed to include a sequence homologous to a region within the RNA of interest as well as a 5' unique sequence that is not present within the RNA of interest. Inclusion of this non-homologous sequence in the RT primer creates a unique tail on the cDNA that is not present within genomic copy of the region of interest. The cDNA is purified by affinity chromatography to remove the RT primer, and then amplified by PCR using a primer complementary to the cDNA and a primer consisting of only the unique sequence of the original RT primer. In this manner, only the cDNA formed in the initial RT reaction can serve as a template during the PCR reaction.

Figure 38. Schematic diagram explaining the principle of ncRT-PCR.

To generate the cDNA an RT primer is designed to include a sequence homologous to the region of interest and a 5' unique sequence (red line) that is not present within the region of interest. The inclusion of this sequence in the RT primer creates a unique tail on the cDNA that is not present within genomic copy of the region of interest. The cDNA (blue line) is column purified to remove the RT primer. The purified cDNA is amplified using a primer complementary to the cDNA (blue arrow) and a primer consisting of only the unique sequence of the original RT primer (red arrow).


5'
3'
RT reaction using primer with non-specific overhang
Affinity purification to remove RT primer
5'
3'
3'
5'
First PCR cycle fills in non-specific region of RT primer
3'
5'
Subsequent PCR cycles use non-specific primer to prevent use of contaminating DNA as a template
5'
3'
3'
5'

To test the ncRT-PCR system a mouse line was utilised which had previously been generated within the lab that contains a single copy of the hCD2 LCR (without the coding sequences of the hCD2 gene) knocked into the CD8 locus between the CD8α and CD8β genes (denoted CD8-LCR knockin. see figure 39A).

Thymic RNA isolated from a CD8-LCR knockin mouse was utilised for ncRT-PCR analysis. The primer combination was designed to specifically detect sense LCR transcripts (figure 36B). In the presence of reverse transcriptase a band of 300bp is generated, indicating the presence of sense coding transcripts in thymocytes from a CD8-LCR knockin mouse (figure 39C). In contrast, only a faint band is observed in RT- lanes. This finding is important, as the RNA samples had not been DNase treated, indicating that ncRT-PCR is an effective means to overcome the problem of genomic DNA contamination. The faint band in the negative controls probably originates from the carry-over of a small fraction of the RT primer during the purification step, allowing amplification from contaminating genomic DNA. Such carry-over could potentially be minimised by increasing the stringency of the wash step during the purification by affinity chromatography. We have yet to apply the ncRT-PCR methodology to single-cell RT-PCR.

7.5 Conclusions

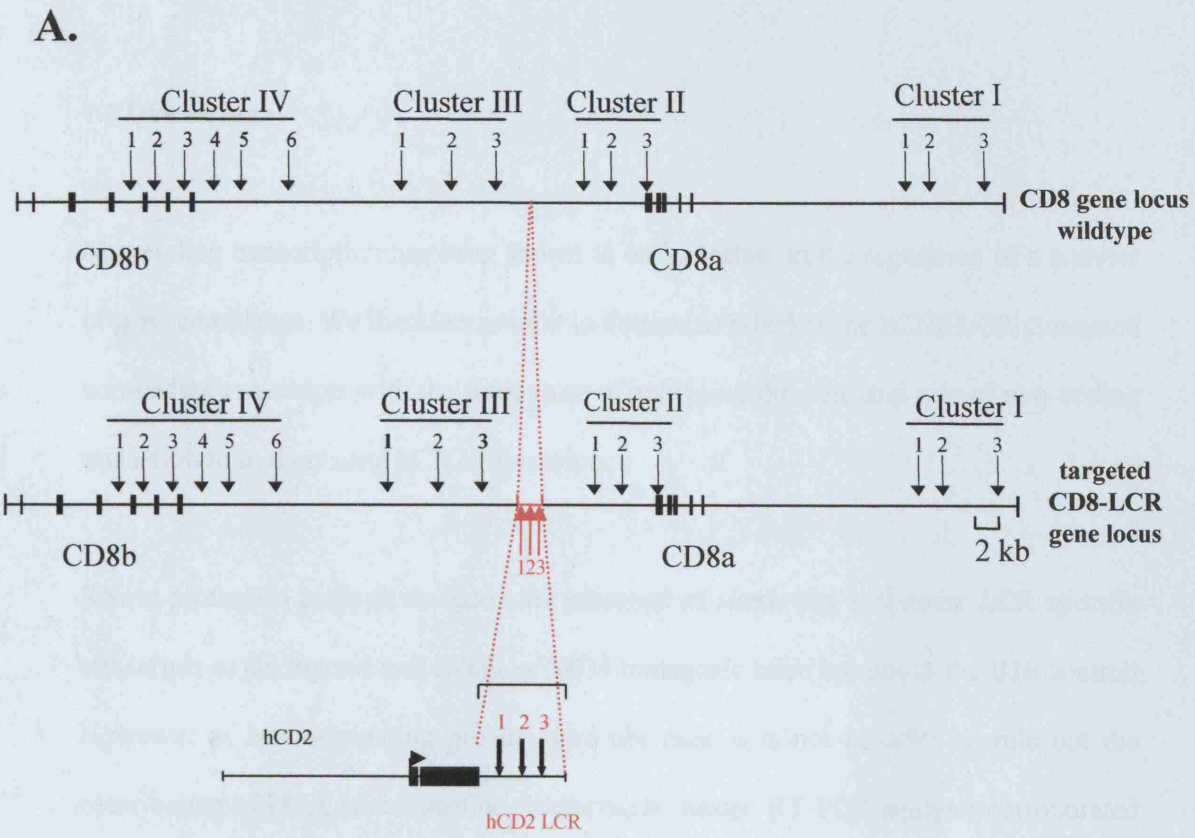
In summary, we have identified the presence of both sense and anti-sense hCD2 LCR transcripts in hCD2 expressing cells isolated from transgenic mice, and sense hCD2 LCR transcript in thymocytes isolated from a CD8-LCR knockin mouse.

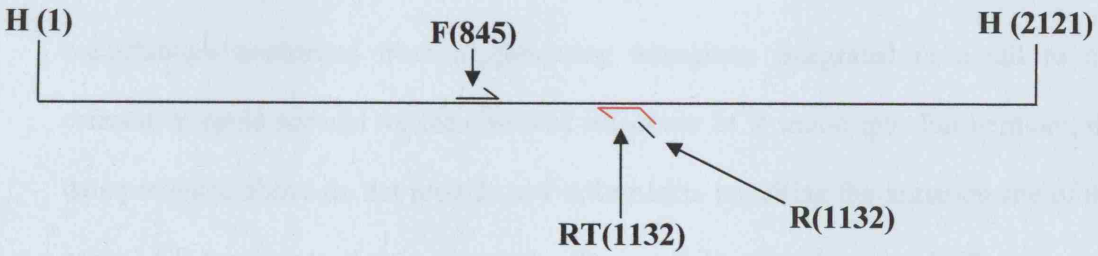
Figure 39. ncRT-PCR analysis to identify sense LCR transcripts in CD8 LCR knockin mice.

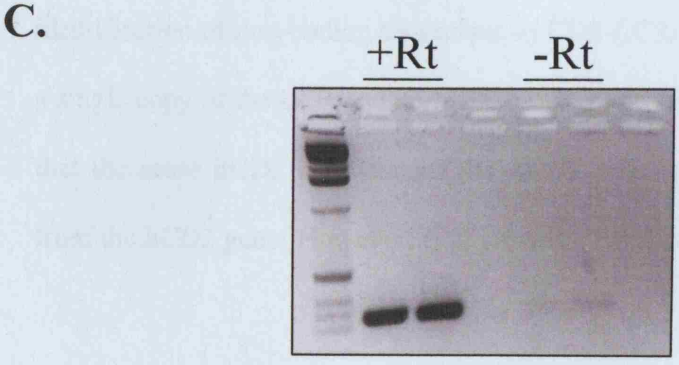
(A) Diagram of the wild type CD8 locus and targeted CD8-LCR locus. Filled boxes represent exons. Below is a diagram of the hCD2 minigene from which the 2.1kb LCR was removed. Hypersensitive sites within the CD8 gene and hCD2 LCR are indicated by arrows.

(B) The 2.1kb hCD2 LCR is shown as demarcated by two HindIII sites (denoted H). The approximate positions of the primers used are indicated (for primer sequences see materials and methods). Numbers in brackets represent the position of the primers in relation to the upstream HindIII demarcating the start of the LCR. RT refers to the primer used for cDNA generation and is shown in red, where as F and R refer to the primers used in the subsequent PCR reaction.

(C) ncRT-PCR analysis was performed on RNA isolated from the thymus of a CD8-LCR knockin mouse. +RT and –RT refers to reactions performed in the presence or absence of reverse transcriptase, respectively. The ncRT-PCR analysis was performed in duplicate.


A.
Cluster IV
Cluster III
Cluster II
Cluster I
1 2 3 4 5 6
1 2 3
1 2 3
1 2 3
CD8 gene locus
wildtype
CD8b
CD8a
Cluster IV
Cluster III
Cluster II
Cluster I
1 2 3 4 5 6
1 2 3
1 2 3
1 2 3
targeted
CD8-LCR
gene locus
CD8b
123
CD8a
2 kb
hCD2
1 2 3
hCD2 LCR


B.
H (1)
F(845)
H (2121)
RT(1132)
R(1132)


C.
+Rt
-Rt

7.6 Discussion

Non-coding transcription has been shown to be important in the regulation of a number of gene complexes. We therefore sought to determine whether the hCD2 LCR generated non-coding transcripts with the future aim of analysing the potential role of non-coding transcription in regulating hCD2 expression.

RNase protection analysis revealed the presence of sense and anti-sense LCR specific transcripts in the thymus and spleen of MG4 transgenic mice but not in the B10 control. However, as intron-spanning probes were not used it is not possible to rule out the contribution of DNA contamination. Importantly, nested RT-PCR analysis corroborated the RNase protection data with the identification of both sense and anti-sense LCR transcripts in the thymus of hCD2.2B transgenic mice, but not within the liver.

One caveat of these experiments is that they both utilised transgenic mice, which contain concatameric arrays of the hCD2 transgene. In such a situation it is possible that read-through transcripts from neighbouring transgenes integrated in a tail to tail orientation could account for the observed anti-sense LCR transcripts. Furthermore, the data presented above do not provide any information regarding the initiation site of the sense LCR transcripts. As a consequence the possibility that the sense LCR transcripts originate as read-through from the hCD2 gene cannot be excluded. However, the identification of non-coding transcripts in CD8-LCR knockin mice, which contain only a single copy of the LCR and no hCD2 coding sequence, gives weight to the possibility that the sense hCD2 LCR transcripts do not arise simply as read through transcripts from the hCD2 gene. However, it is possible that the sense LCR transcripts detected in

the CD8-LCR knockin mice are generated by naturally occurring non-coding CD8 locus transcripts that read-through the LCR.

7.6.1 Why is the hCD2 LCR Transcribed in Both Sense and Anti-Sense Directions?

It seems paradoxical that both sense and anti-sense transcripts are produced simultaneously from the same sequence within the hCD2 LCR, as RNA Pol II molecules transcribing in opposing orientations would collide resulting in pausing and ultimately transcriptional termination. Furthermore, the presence of both sense and anti-sense transcripts would be expected to form double stranded RNA resulting in the activation of the RNAi machinery. It therefore seems unlikely that these transcripts are produced simultaneously within a single cell. Interestingly, within the *Igh* locus both sense and anti-sense non-coding transcripts have been identified [496]. However, both sense and anti-sense transcripts were rarely found in the same cell [496].

It is therefore important to note that all the analysis reported herein was performed on RNA isolated from bulk populations of cells. It is therefore possible that the LCR is only transcribed in one direction in any cell, but the direction of transcription is heterogeneous within a population of cells. It is also possible that the transcripts detected only occur in a few cells within the population. To address this possibility we have recently attempted to utilise multiplex single cell RT-PCR (data not shown). However, in this system we have been unable to overcome the problem of DNA contamination. The ncRT-PCR technology may overcome this problem.

Alternatively, all cells may express both transcripts but the direction of transcription across the LCR is alternated. In addition, the type of transcript generated may be

regulated in a cell cycle dependent manner. This would be similar to regulation of β-globin intergenic transcription. The β-globin LCR and intergenic transcripts are generated almost exclusively within G1 of the cell cycle [390]. Repeating the RT-PCR analysis on sorted populations of G1, S, and G2 cells may reveal whether hCD2 LCR transcription is regulated in a cell-cycle dependent manner.

Although both sense and anti-sense LCR specific transcripts have been identified, only transgenic systems have been analysed and not human cells, raising the possibility that they may result from a transgenic artefact. To determine whether the endogenous hCD2 LCR produces such transcripts, ncRT-PCR analysis will be performed on RNA isolated from human cell lines. Preliminary data (not shown) indicates that the hCD2 LCR is transcribed in Jurkat cells (a human cell line derived from a T cell leukaemia).

Furthermore, the data reported herein does not allow determination of an *in vivo* function of these transcripts. It is possible that these transcripts are non-functional, originating by chance through illegitimate transcription from pseudo promoters within the open chromatin of the hCD2 LCR in T cells. Indeed, there is an Alu repeat within the LCR in the same transcriptional orientation as the hCD2 gene (spanning the region 898-1226, relative to the upstream HinDIII site which demarcates the start of the 2.1kb LCR). This is potentially relevant as Alu repeats are known to contain a functional Pol III promoter [500]. Furthermore, the probe in the RNase protection assay lies within the Alu repeat. Importantly, however, the observation that transcripts were not detected in non-transgenic mice confirms that the RNA product detected in these assays is generated specifically from the hCD2 LCR and not from additional Alu repeats within the murine genome. In contrast, the 5' PCR primer utilised in the RT-PCR analysis

originates from upstream of the Alu repeat, indicating that sense non-coding LCR transcripts originate from upstream of the Alu repeat, although this does not rule out the possibility that the Alu Pol III promoter is active. In addition, the hCD2 LCR anti-sense transcripts are transcribed in the opposite orientation to the Alu element and therefore cannot have originated from this repeat.

To establish a possible function for the non-coding transcripts rapid amplification of cDNA ends (RACE) will be performed on RNA isolated from hCD2 transgenic thymocytes and Jurkat cells, to determine the start site of each transcript. If the transcriptional start sites of either the sense or anti-sense transcripts are internal to the LCR, modified hCD2 constructs will be generated in which the start sites have been independently deleted. Transgenic mice containing these constructs will be analysed by ncRT-PCR to determine whether the deletion has affected the generation of the sense or anti-sense LCR transcripts. Furthermore, flow cytometry will be used to determine the effect, if any, on hCD2 expression.

Chapter Eight

Methylation of the hCD2 LCR

8.1 Gene Regulation and DNA Methylation

Genes are frequently methylated at the DNA level in tissues in which they are not expressed. It is thought that DNA methylation causes gene silencing and chromatin condensation thus preventing non-specific gene expression [501]. During development, genes expressed in a tissue specific manner become selectively demethylated in the appropriate cell types to allow gene expression [502, 503].

A well-characterised example of such tissue specific demethylation is that of the T cell receptor (TCR) α/δ locus [400]. Expression from the TCRα/δ locus is strictly regulated by an LCR that lies downstream of the TCRα gene [400, 504, 505]. The TCRα LCR contains nine DNase I HSSs and is able to confer high-level, position-independent, copy number-dependent and T cell specific expression to linked transgenes in mice [400, 504, 505]. In non-expressing tissues the LCR is heavily methylated and the locus is silenced [400]. To allow initiation of gene expression the LCR is specifically demethylated in a tissue specific manner [400]. This tissue specific demethylation process is driven by two HSSs within the TCRα LCR [400]. Deletion of these HSSs results in the loss of LCR demethylation and a concomitant reduction in gene expression [400]. Interestingly the region of the TCRα LCR that is specifically demethylated coincides with a HSS, although the significance of this finding is not known [400].

The previous section described the identification of tissue-specific non-coding transcripts within the hCD2 LCR. Methylation of the LCR could potentially provide a mechanism to regulate the generation these non-coding transcripts. Alternatively, methylation of the LCR could directly influence expression of the hCD2 gene. It was

therefore interesting to determine whether sequences (especially the DNase I HSSs) within the hCD2 LCR are differentially methylated within expressing and non-expressing tissues.

To examine the DNA methylation patterns in expressing and non-expressing tissues an assay using the methylation sensitive restriction enzyme McrBC was utilised. The assay, which was developed by Santoso B et al. is similar to a DNase I HSS assay except that a titration of McrBC is used in place of DNase I [400].

McrBC is a methylation dependent restriction enzyme that recognises a pair of methylated cytosine residues within the sequence 5'-PumC(N40-20000)PumC-3' and cleaves within 30bp from one of the methylated residues [400]. Titration of this enzyme onto genomic DNA allows the scanning of genomic DNA without the requirement for particular restriction enzyme recognition sites within the region of interest [400]. McrBC treatment of isolated genomic DNA in a titration assay generates different sized fragments that can be resolved and detected by Southern blot hybridisation [400]. This allows the gross mapping of the CpG methylation status of a given locus.

8.2 Mapping the Gross Methylation Status of the hCD2 LCR in hCD2 Transgenic Mice and Human Cell Lines

Genomic DNA was extracted from thymus, spleen and kidney tissue from an MG4 hCD2 transgenic mouse. Genomic DNA was also isolated from HEK 293 cells (a cell line derived from human embryonic kidney cells that does not express hCD2) and Jurkat cells (a cell line derived from a human T cell leukaemia that expresses hCD2).

DNA was digested with HinDIII before digestion with increasing amounts of McrBC (see materials and methods). The reaction was stopped after 20 minutes by the addition of a stop solution (see materials and methods). The McrBC treated DNA samples were run on a 1% agarose gel and subjected to Southern blot analysis. The resulting blot was probed with a 5'LCR probe (figure 40A). A 2kb parent band corresponding to the HinDIII LCR fragment is evident in all lanes. It must be noted that although equal quantities of DNA were loaded into each lane the parent band is much more intense in the lanes containing the MG4 genomic DNA. This is because the homozygous MG4 contains 12 copies of the hCD2 transgene, whereas the human cell lines contain only 2 copies of the endogenous gene. A second band of 900bp, which most probably corresponds to an end fragment from an incomplete copy of the hCD2 transgene, is also evident in all lanes from the MG4. The presence of any additional bands indicates the presence of methylated sites within the LCR. In the lanes corresponding to DNA isolated from MG4 kidney cells and HEK 293 cells a band of approximately 1.1kb is generated on digestion with increasing levels of McrBC. This band is not evident in the lanes corresponding to DNA isolated from the MG4 thymus or spleen cells or from Jurkat cells even with the highest level of McrBC digestion. Interestingly, on longer exposures the 1.1kb methylation band becomes apparent in the spleen samples (data not shown). This is can be explained by the fact that the spleen contains a mixed population of cells, not all of which express hCD2.

These results show that hCD2 LCR is differentially methylated in non-expressing and expressing tissues, in both transgenic mice and human cell lines.

8.3 Comparing the Relative position of the hCD2 LCR Methylation site to hCD2 LCR specific DNase I HSSs

Genomic DNA was isolated from heart and kidney tissue taken from an MG4 transgenic mouse. DNA was digested with HinDIII before digestion with increasing amounts of McrBC. The reaction was stopped after 20 minutes by the addition of a stop solution. In parallel, nuclei extracted from MG4 thymocytes were treated with increasing amounts of DNase I. Genomic DNA isolated from DNase I treated nuclei was digested with HinDIII to allow resolution of the HSSs within the hCD2 LCR. The McrBC and DNase I treated DNA samples were run in parallel on a 1% agarose gel and subjected to Southern blot analysis. The resulting blot was probed with a 5'LCR probe.

Figure 40B shows the methylation sites and DNase I HSSs of the 2kb LCR in non-expressing and expressing tissues, respectively. A 2kb parent band corresponding to the LCR HinDIII fragment is evident in all lanes. A second band of 900bp corresponding to an end fragment from an incomplete copy of the hCD2 transgene is also evident in all lanes. As previously noted treatment of genomic DNA isolated from tissues that do not express hCD2 (in this case heart and kidney) with increasing levels of McrBC results in the generation of an intense band of 1.1kb. An additional weak band of ~1.6kb is evident in the McrBC treated genomic DNA isolated from heart tissue. Running DNase I treated genomic DNA extracted from MG4 thymus on the same gel allows the relative positioning of the methylation sites and DNaseI HSS within the LCR. Treatment of MG4 thymic nuclei with increasing levels of DNase I results in the production two strong HSSs of size 500-600bp and 1.6kb corresponding to HSS1 and HSS3, respectively [341]. A weaker HSS of 1.1-1.3kb has also been reported and corresponds to HSS2 [341]. Comparing the McrBC and DNase I treated samples reveals that HSS1

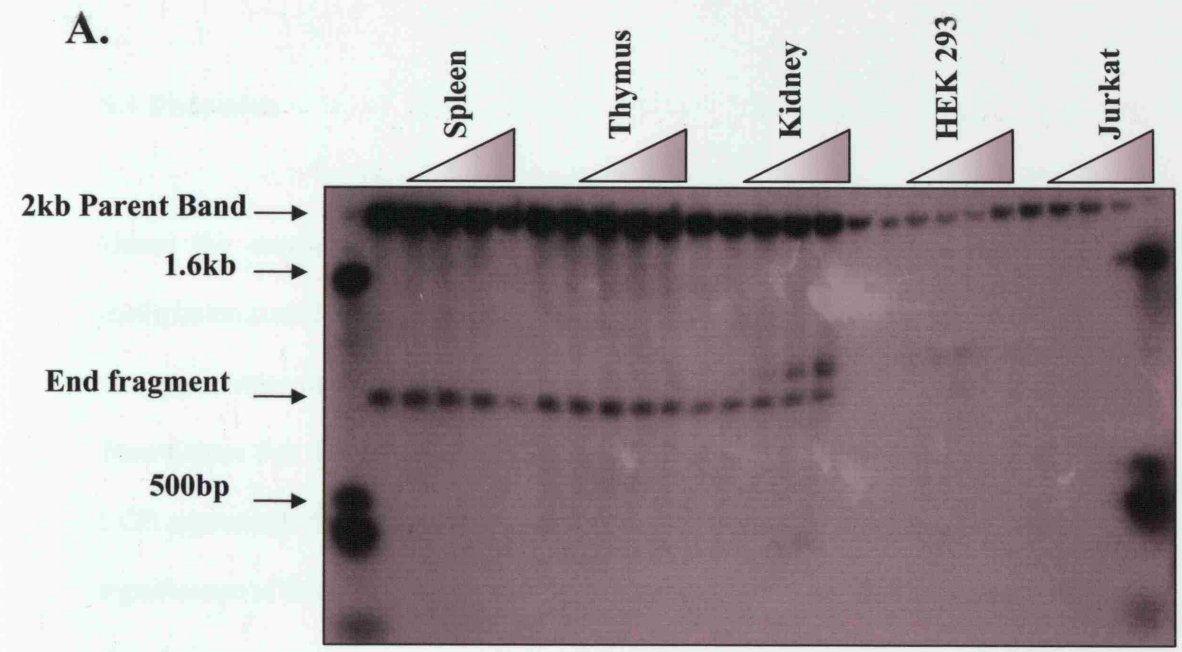
does not correlate with a region of detectable methylation in any of the tissues analysed. In contrast, the region known to contain HSS2 correlates approximately with the strong methylation site in non-expressing tissues. Finally, HSS3 appears to correlate approximately with a weakly methylated region found in genomic DNA isolated from MG4 heart. However, this experiment needs to be repeated to confirm that this is a true methylation site in heart tissue. Figure 40C shows a diagram indicating the relative position of the hCD2 LCR HSSs and methylation sites.

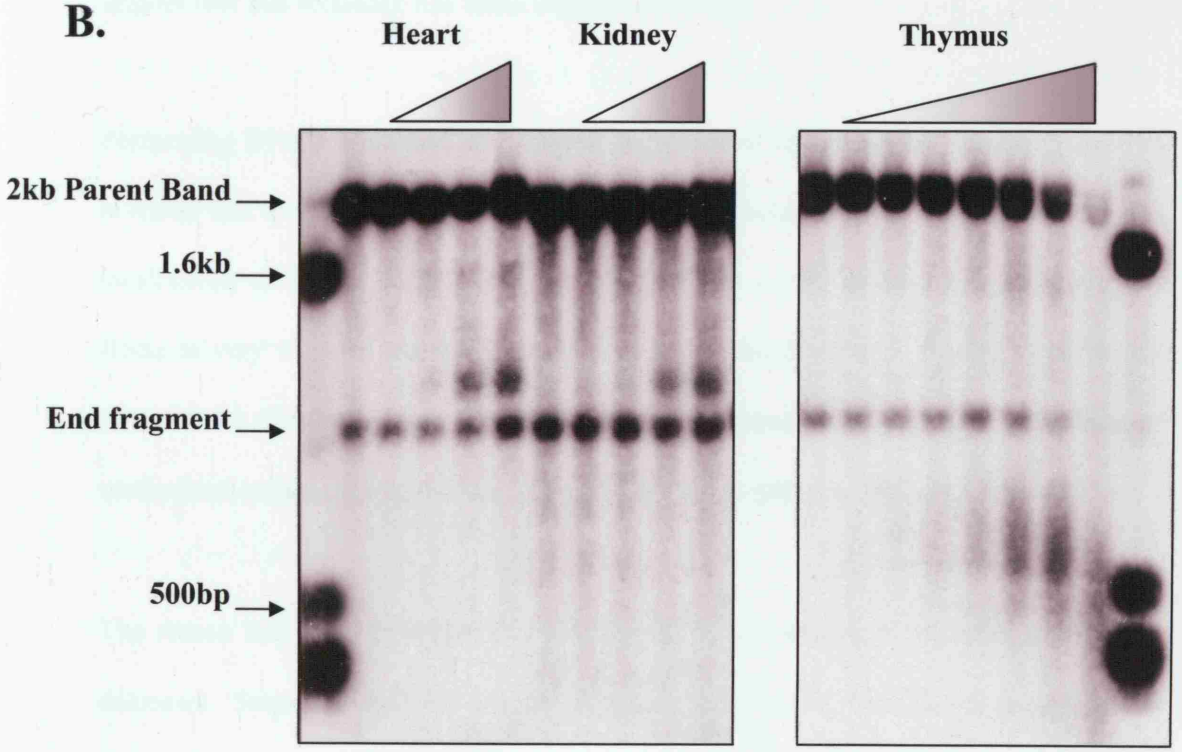
Figure 40. Gross mapping of DNA methylation sites within the hCD2 LCR.

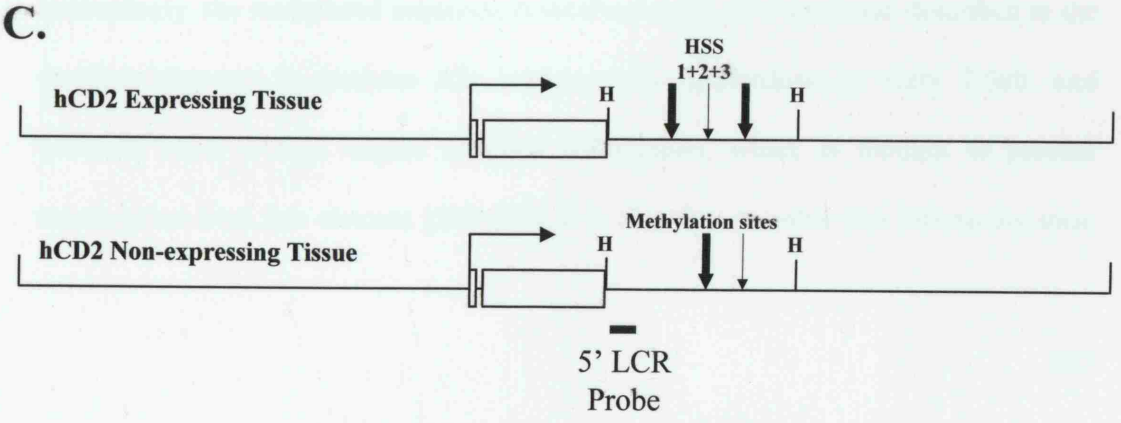
(A) Southern blot showing methylation mapping of the hCD2 LCR. DNA from spleen, thymus and kidney tissue taken from an MG4 mouse as well as DNA from HEK293 cells and Jurkat cells was digested with HinDIII and then treated with increasing amounts of McrBC. The location of the 2kb parent band and 1kb end fragment are indicated.

(B) Southern blot showing methylation mapping and DNase I analysis of the hCD2 LCR. DNA from heart and kidney tissue taken from an MG4 mouse was digested with HinDIII and then treated with increasing amounts of McrBC. In parallel MG4 thymocyte nuclei were treated with increasing amounts of DNase I, before digestion with HinDIII. The location of the 2kb parent band and 1kb end fragment are indicated.

(C) Schematic diagram of the hCD2 minigene construct which was used to generate the MG4 transgenic line. In a tissue expressing hCD2 three hypersensitive sites (HSSs) are present within the hCD2 LCR, as shown by three arrows. The thickness of the arrows represents the relative strength of each HSS. In a tissue that does not express hCD2 no HSSs are present within the hCD2 LCR. However, the hCD2 LCR is methylated. The positions of the methylation sites in the hCD2 LCR are indicated. The thickness of the arrows represents the relative strength of the methylation sites. The positions of the HinDIII restriction sites and 5'LCR probe also are indicated. Open boxes represent exons. H=HinDIII.


A.
Spleen
Thymus
Kidney
HEK 293
Jurkat
2kb Parent Band
1.6kb
End fragment
500bp


B.
Heart
Kidney
Thymus
2kb Parent Band
1.6kb
End fragment
500bp


C.
HSS
1+2+3
hCD2 Expressing Tissue
H
H
Methylation sites
hCD2 Non-expressing Tissue
H
H
5' LCR
Probe

8.4 Discussion

Using the methylation sensitive enzyme McrBC we have analysed the gross methylation status of the hCD2 LCR in expressing and non-expressing tissues of hCD2 transgenic mice and in human cells lines. In both transgenic mice and humans, cells from tissues that do not express hCD2 show a strong methylation site within the hCD2 LCR approximately 1.1kb downstream of the polyadenylation signal. The physiological significance of this methylation site has not been determined. However, the observation that the same sequence was methylated in both transgenic mice and a human cell line implies that this sequence has some unique properties.

Performing DNase I analysis of thymocytes from a hCD2 transgenic mouse in parallel revealed that the strong methylation site in non-expressing tissues is correlated with the localisation of DNase I HSS2 in thymocytes. This result was initially surprising as HSS2 is very weak in transgenic mice. However, this finding is similar to the result obtained by Santoso et al. who revealed a correlation between the localisation of methylated sequences and a weak DNase I HSS site within the TCRα LCR [400].

The reason that this site within the hCD2 LCR is specifically methylated is currently unknown. Sequence analysis reveals that the hCD2 LCR methylated region is not unusually rich in CpG dinucleotides when compared to the rest of the LCR sequence. Interestingly, the methylated sequence is localised to an Alu repeat (as described in the previous chapter). In humans Alu repeats occur approximately every 2.5kb, and generally show a high degree of CpG methylation, which is thought to prevent transcription from this element [506-508]. It is therefore possible that this methylation

site simply corresponds to a silenced Alu repeat. However, this suggestion does not readily explain why the site becomes demethylated in tissues that express hCD2. It is potentially interesting that methylation of the LCR and the generation of LCR non-coding transcripts are inversely correlated.

Previously, an Alu repeat located within the human keratin 18 LCR has been reported to direct copy number-dependant, position-independent expression in transgenic mice in the absence of the other LCR sequences [509]. Conversely, a point mutation within the Alu repeat RNA Pol III promoter abolishes the ability of the human keratin 18 LCR to confer copy number-dependant expression, although position independent expression is not affected [510]. This provides the intriguing possibility that the Alu repeat within the hCD2 LCR is important for directing hCD2 expression.

Importantly, we have not determined whether methylation of this region within the LCR is repressive for hCD2 gene transcription or the generation LCR derived non-coding transcripts. Furthermore, we do not know if the LCR is actively demethylated to allow hCD2 expression, or whether demethylation occurs as a secondary effect resulting from active transcription within the hCD2 coding region or LCR sequences. Regulatory elements that are able to direct tissue specific demethylation have been characterised within the TCRα LCR and the *Igκ* locus [400, 511]. It would therefore be interesting to determine whether such an element exists within the hCD2 LCR. One potential candidate for this demethylation element is HSS3. Performing the methylation assay on sorted populations of $hCD2^+$ and $hCD2^-$ thymocytes from variegating hCD2-1.3kb LCR transgenic mice may reveal a differential methylation associated with the variegated phenotype.

Chapter Nine

Final Discussion

9.1 Project aim

The overall aim of this project was to study the control of hCD2 gene expression and how different parts of the gene contribute to its regulation. This was achieved by searching for novel regulatory elements within the hCD2 gene whose function could be addressed by deleting them using an *in vivo* Cre/loxP deletion strategy. In order to fulfil this aim it was necessary to develop tools that would enable the *in vivo* tissue-specific deletion of these identified sequences utilising the Cre/loxP based approach.

9.2 Identification of candidate regulatory elements for tissue-specific Cre/loxP deletion

Screening the hCD2 gene using a range of different techniques identified a number of novel potential regulatory elements.

Cre/loxP deletion strategies require the use of single copy transgenic lines, however, expression from a single copy of the hCD2 minigene falls below the detection level of flow cytometry. Therefore, it was necessary to determine whether there were any elements residing within the full hCD2 gene that when included in the minigene would impart elevated levels of expression. In addition, as the Cre/loxP deletion approach will be used in variegation studies it is important that inclusion of these additional sequences does not render the transgene insensitive to PEV. Previously, intron 4 had been shown to confer elevated levels of expression on hCD2 transgenes, however, intron 4 had never been tested in context of variegating hCD2 transgenic lines. Thus, it was possible that inclusion of intron 4 could compensate for the absence of HSS3, preventing its use in variegation studies. To investigate this, transgenic mice carrying a hCD2 construct that contained intron 4, the disabled 1.3kb LCR and a single loxP site at the terminal 3'

end of the transgene (denoted 1.3loxP) were generated. Analysis of hCD2 expression in these mice revealed two important findings. First, inclusion of intron 4 did not prevent variegation of hCD2 transgenes containing a truncated LCR, whilst allowing elevated levels of expression. Second, using a Cre/loxP based approach to address the influence of transgene copy number on hCD2 variegation it was shown that decreasing transgene copy number whilst retaining the same site of integration affected hCD2 variegation in a manner that was dependant upon the site of integration. Thus, in two transgenic lines copy number reduction resulted in an increase in the level of variegation whereas, in one case copy number reduction resulted in an decrease in variegation. Finally, two transgenic lines showed no change in the level of variegation on copy number reduction. These results confirm that only single copy transgenes can be utilised in Cre/loxP deletion experiments designed to address variegation issues.

Interestingly, imprinted expression was noted in a number of the 1.3loxP lines. For example, in mice in which the transgene was of the maternal origin imprinting was evident as a decrease in the level of hCD2 expression and an increase in level of PEV, when compared to mice in which the transgene was of the paternal origin. The mechanism responsible for mediating this imprinting was not addressed. Yet, as hCD2 has never shown imprinted expression before, it is unlikely that this finding represents a normal mode of hCD2 regulation. It is more likely that this imprinting reflects a general differential epigenetic modification of the germline DNA between developing oocytes and spermatocytes. It remains unclear for the moment why the 1.3loxP transgene should be so susceptible to this differential epigenetic modification. This result is particularly important as it highlights the need to always consider the parental origin of the transgene, even when comparing mice in which the transgene is in the same site of

integration. Previously, non-imprinted genes have been noted to display imprinted expression when present as transgenes. However, the mechanisms responsible for directing this imprinting have never been identified.

S/MAR sites have been shown to have important roles in a diverse range of biological events, such as sister chromatid separation and the regulation of gene expression. Indeed, a number of S/MAR sites have been shown to protect linked transgenes against PEV. It was therefore interesting to examine the hCD2 gene for the presence of potential S/MAR sites. Using a bioinformatics approach a potential S/MAR site was identified within the 5' region of the hCD2 gene. The presence of DNase I hypersensitive sites normally indicates the position of regulatory elements, including S/MARS. Therefore, DNase I hypersensitivity analysis was used to screen the region containing the predicted S/MAR. This analysis revealed that the predicted S/MAR site did not form a DNase I HSS in transgenic mice. Furthermore, transgenic mice in which the predicted S/MAR site had been removed revealed that the predicted S/MAR site was not essential for directing hCD2 gene expression. Taken together, these two findings suggest that the predicted S/MAR site does not play an essential role in the regulation of hCD2 expression, however, they do not rule out the possibility that these sequences play a role in regulating hCD2 expression when the gene is present in its natural chromosomal site. Indeed, the sequences containing the predicted S/MAR site are conserved in the human, mouse, rat and bovine genome, suggesting a functional role for this region. As the S/MAR site is contained within a specified region it is an attractive candidate for tissue-specific Cre/loxP deletion studies. This would allow the direct comparison of transgenes at the same site of integration in which the predicted S/MAR element was present or absent. However, it is possible that the conserved

S/MAR site only functions at the endogenous site and not at transgenic loci. Thus, a more informative approach would be to delete the conserved region containing the S/MAR site from the endogenous mouse CD2 gene and analyse the effect on mouse CD2 expression.

Previously, intergenic transcription has been implicated in the regulation of gene expression. For example, transcription of the β-globin LCR and intergenic regions is thought to be important for the regulated expression of the globin genes. Therefore, the potential role of non-coding transcription in the regulation of hCD2 expression was investigated. Using RNase protection analysis and nested RT-PCR both sense and anti-sense transcripts originating from the hCD2 LCR were identified in transgenic mice. To investigate a potential contribution of non-coding transcription to regulation of hCD2 expression it would be interesting to abolish non-coding transcription within the LCR. However, as the origin or promoter of these transcripts has yet to be determined it is not possible to address the role of these transcripts utilising a Cre/loxP based approach. Due to time constraints further analysis of these transcripts could not undertaken before submission of this thesis.

Finally, DNA methylation has been shown to be an important negative regulator of gene expression. It was therefore interesting to examine the hCD2 LCR for the presence of methylated sequences. Using the DNA methylation sensitive enzyme McrBC, a differentially methylated site was identified with the hCD2 LCR, the position of which correlated with the presence of a previously identified DNase I HSS, HSS2. This differentially methylated site is methylated in tissues that do not express hCD2 and unmethylated in tissues that express hCD2. It is already known that the sequences

proximal to this methylation site (HSS2) are essential for preventing PEV in transgenic mice. However, like HSS3 it is not known whether HSS2 is required only to initiate protection against PEV or whether it is required to maintain protection against PEV. Thus, this element represents a suitable candidate for Cre/loxP deletion analysis. In addition, it would be interesting to scan the rest of the hCD2 gene using the same methylation sensitive enzyme McrBC. This may lead to the identification of additional regulatory elements within the hCD2 gene. Furthermore, as DNA methylation is considered to be one of the main regulators of imprinting it would be particularly interesting to look in 1.3loxP mice for differentially methylated regions within the transgene.

9.3 The development of tools for mediating tissue-specific Cre/loxP deletions of regulatory elements from the hCD2 gene in vivo

To enable the tissue-specific deletion of floxed regulatory elements from the hCD2 gene two different Cre expressing transgenic lines were generated. The hCD2-iCre and Vav-iCre lines utilise the hCD2 and Vav regulatory elements to drive the tissue-specific expression of the iCre gene, respectively. These different lines were tested utilising two complementary approaches. The first approach utilised the ROSA26-EYFP reporter line to enable positive identification of the cell types in which the Cre transgene was expressed. Analysis revealed that the hCD2-iCre line mediated Cre recombination in virtually all lymphoid cells, whereas, the Vav-iCre line mediated recombination in all haematopoietic cell types analysed.

The second approach utilised the Cre-inducible diphtheria toxin line, DIP-R1, to reveal the true recombination efficiency in each Cre expressing line. The hCD2-iCre line was

able to drive ablation specifically in T cells and B cells, although ablation was not 100% efficient. Southern blot analysis revealed that surviving lymphoid cells in hCD2-iCre/DIP-R1 mice resulted from inefficiencies in Cre mediated deletion and not from a lack of diphtheria toxin toxicity. Interestingly, no Vav-iCre/DIP-R1 mice were obtained for analysis suggesting that the Cre-mediated ablation throughout the entire haematopoietic compartment (as predicted for the Vav-iCre line) is efficient enough in this system to lead to embryonic death.

In combination, these results indicate that the hCD2-iCre transgene does not drive recombination in 100% of lymphoid cells. However, this inefficiency only becomes apparent when cells in which recombination has failed to occur have a selective advantage. This finding may be relevant in the generation of tissue-specific gene knockouts that influence the fate of the mutant cells. However, as deletion of regulatory elements from hCD2 transgenes has no impact on the fate of cells in which recombination has occurred both the hCD2-iCre and Vav-iCre lines will be valuable in directing tissue-specific deletions of hCD2 regulatory elements. Currently, a line that drives expression of a progesterone-inducible Cre is under development in the lab of Alexandre Potocnik in the division of molecular immunology at the NIMR. This Cre line will be useful in mediating inducible deletion of regulatory elements from hCD2 transgenes at different stages of T cell development.

9.4 Conclusions

In summary, inclusion of intron 4 in hCD2 constructs allows increased levels of expression without conferring resistance to PEV. Therefore, transgenes containing intron 4 will form the basis for all Cre/loxP based deletion studies that will be undertaken in the future. To determine the function of novel elements identified with the hCD2 gene, loxP sites will be placed either side of each region of interest. Transgenic mice containing such transgenes will be crossed to the two Cre expressing transgenic lines that have been developed, resulting in the tissue-specific deletion of the floxed target sequence. This will allow the direct comparison of transgenes at the same site of integration in which the target sequence is present or absent. Furthermore, the use of the two different Cre expressing lines will allow the target sequence to be deleted at different developmental time points.

Chapter Ten

References

1 **Tsukiyama, T.,** The in vivo functions of ATP-dependent chromatin-remodelling factors. *Nat Rev Mol Cell Biol* 2002. **3**: 422-429.
2 **Donaldson, K. M., Lui, A. and Karpen, G. H.,** Modifiers of terminal deficiency-associated position effect variegation in Drosophila. *Genetics* 2002. **160**: 995-1009.
3 **Wakimoto, B. T.,** Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell* 1998. **93**: 321-324.
4 **Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J.,** Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 1997. **389**: 251-260.
5 **Felsenfeld, G. and Groudine, M.,** Controlling the double helix. *Nature* 2003. **421**: 448-453.
6 **Horn, P. J. and Peterson, C. L.,** Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* 2002. **297**: 1824-1827.
7 **Hansen, J. C.,** Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct* 2002. **31**: 361-392.
8 **Peterson, C. L. and Laniel, M. A.,** Histones and histone modifications. *Curr Biol* 2004. **14**: R546-551.
9 **Leuba, S. H., Yang, G., Robert, C., Samori, B., van Holde, K., Zlatanova, J. and Bustamante, C.,** Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. *Proc Natl Acad Sci U S A* 1994. **91**: 11621-11625.
10 **Parseghian, M. H. and Hamkalo, B. A.,** A compendium of the histone H1 family of somatic subtypes: an elusive cast of characters and their characteristics. *Biochem Cell Biol* 2001. **79**: 289-304.
11 **Horn, P. J. and Peterson, C. L.,** The bromodomain: a regulator of ATP-dependent chromatin remodeling? *Front Biosci* 2001. **6**: D1019-1023.
12 **Adkins, N. L., Watts, M. and Georgel, P. T.,** To the 30-nm chromatin fiber and beyond. *Biochim Biophys Acta* 2004. **1677**: 12-23.
13 **Belmont, A. S., Braunfeld, M. B., Sedat, J. W. and Agard, D. A.,** Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. *Chromosoma* 1989. **98**: 129-143.
14 **Belmont, A. S. and Bruce, K.,** Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol* 1994. **127**: 287-302.
15 **Tumbar, T., Sudlow, G. and Belmont, A. S.,** Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol* 1999. **145**: 1341-1354.
16 **Muller, W. G., Walker, D., Hager, G. L. and McNally, J. G.,** Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J Cell Biol* 2001. **154**: 33-48.
17 **Lorch, Y., LaPointe, J. W. and Kornberg, R. D.,** Initiation on chromatin templates in a yeast RNA polymerase II transcription system. *Genes Dev* 1992. **6**: 2282-2287.

18 **Workman, J. L. and Roeder, R. G.,** Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* 1987. **51**: 613-622.

19 **Izban, M. G. and Luse, D. S.,** Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev* 1991. **5**: 683-696.

20 **Boeger, H., Bushnell, D. A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J. S., Westover, K. D. and Kornberg, R. D.,** Structural basis of eukaryotic gene transcription. *FEBS Lett* 2005. **579**: 899-903.

21 **Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S.,** Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 2002. **9**: 279-289.

22 **Berger, S. L.,** Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 2002. **12**: 142-148.

23 **Cosgrove, M. S., Boeke, J. D. and Wolberger, C.,** Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* 2004. **11**: 1037-1043.

24 **Fischle, W., Wang, Y. and Allis, C. D.,** Histone and chromatin cross-talk. *Curr Opin Cell Biol* 2003. **15**: 172-183.

25 **Jenuwein, T. and Allis, C. D.,** Translating the histone code. *Science* 2001. **293**: 1074-1080.

26 **Kouskouti, A. and Talianidis, I.,** Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *Embo J* 2005. **24**: 347-357.

27 **Miao, F. and Natarajan, R.,** Mapping global histone methylation patterns in the coding regions of human genes. *Mol Cell Biol* 2005. **25**: 4650-4661.

28 **Liang, G., Lin, J. C., Wei, V., Yoo, C., Cheng, J. C., Nguyen, C. T., Weisenberger, D. J., Egger, G., Takai, D., Gonzales, F. A. and Jones, P. A.,** Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A* 2004. **101**: 7357-7362.

29 **Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L. and Lander, E. S.,** Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 2005. **120**: 169-181.

30 **Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P. and Groudine, M.,** The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 2004. **18**: 1263-1271.

31 **Roh, T. Y., Cuddapah, S. and Zhao, K.,** Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 2005. **19**: 542-552.

32 **Kurdistani, S. K., Tavazoie, S. and Grunstein, M.,** Mapping global histone acetylation patterns to gene expression. *Cell* 2004. **117**: 721-733.

33 **Dion, M. F., Altschuler, S. J., Wu, L. F. and Rando, O. J.,** Genomic characterization reveals a simple histone H4 acetylation code. *Proc Natl Acad Sci U S A* 2005. **102**: 5501-5506.

34 **Kim, A. and Dean, A.,** Developmental stage differences in chromatin subdomains of the beta-globin locus. *Proc Natl Acad Sci U S A* 2004. **101**: 7028-7033.

35 **Wang, L., Kametani, Y., Katano, I. and Habu, S.,** T-cell specific enhancement of histone H3 acetylation in 5' flanking region of the IL-2 gene. *Biochem Biophys Res Commun* 2005. **331**: 589-594.

36 **Wolffe, A. P. and Hayes, J. J.,** Chromatin disruption and modification. *Nucleic Acids Res* 1999. **27**: 711-720.

37 **Hebbes, T. R., Clayton, A. L., Thorne, A. W. and Crane-Robinson, C.,** Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *Embo J* 1994. **13**: 1823-1830.

38 **Sengupta, N. and Seto, E.,** Regulation of histone deacetylase activities. *J Cell Biochem* 2004. **93**: 57-67.

39 **Katan-Khaykovich, Y. and Struhl, K.,** Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev* 2002. **16**: 743-752.

40 **Ozdemir, A., Spicuglia, S., Lasonder, E., Vermeulen, M., Campsteijn, C., Stunnenberg, H. G. and Logie, C.,** Characterization of lysine 56 of histone H3 as an acetylation site in Saccharomyces cerevisiae. *J Biol Chem* 2005. **280**: 25949-25952.

41 **Ye, J., Ai, X., Eugeni, E. E., Zhang, L., Carpenter, L. R., Jelinek, M. A., Freitas, M. A. and Parthun, M. R.,** Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. *Mol Cell* 2005. **18**: 123-130.

42 **Noma, K., Allis, C. D. and Grewal, S. I.,** Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 2001. **293**: 1150-1155.

43 **Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. and Felsenfeld, G.,** Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 2001. **293**: 2453-2455.

44 **Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T.,** Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000. **406**: 593-599.

45 **Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. and Grewal, S. I.,** Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 2001. **292**: 110-113.

46 **Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. and Kouzarides, T.,** Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 2001. **410**: 120-124.

47 **Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T.,** Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 2001. **410**: 116-120.

48 **Delattre, M., Spierer, A., Tonka, C. H. and Spierer, P.,** The genomic silencing of position-effect variegation in Drosophila melanogaster: interaction between the heterochromatin-associated proteins Su(var)3-7 and HP1. *J Cell Sci* 2000. **113 Pt 23**: 4253-4261.

49 **Schotta, G., Ebert, A., Dorn, R. and Reuter, G.,** Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Semin Cell Dev Biol* 2003. **14**: 67-75.

50 **Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S. and Zhang, Y.,** Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 2002. **298**: 1039-1043.

51 **Stewart, M. D., Li, J. and Wong, J.,** Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. *Mol Cell Biol* 2005. **25**: 2525-2538.

52 **Ng, H. H., Robert, F., Young, R. A. and Struhl, K.,** Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 2003. **11**: 709-719.

53 **Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. and Reinberg, D.,** Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev* 2002. **16**: 479-489.

54 **Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P. and Zhang, Y.,** Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol Cell* 2001. **8**: 1207-1217.

55 **Zegerman, P., Canas, B., Pappin, D. and Kouzarides, T.,** Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J Biol Chem* 2002. **277**: 11621-11624.

56 **Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. and Kouzarides, T.,** Active genes are tri-methylated at K4 of histone H3. *Nature* 2002. **419**: 407-411.

57 **Kubicek, S. and Jenuwein, T.,** A crack in histone lysine methylation. *Cell* 2004. **119**: 903-906.

58 **Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A. and Casero, R. A.,** Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 2004. **119**: 941-953.

59 **Trewick, S. C., McLaughlin, P. J. and Allshire, R. C.,** Methylation: lost in hydroxylation? *EMBO Rep* 2005. **6**: 315-320.

60 **Nowak, S. J. and Corces, V. G.,** Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet* 2004. **20**: 214-220.

61 **Wei, Y., Mizzen, C. A., Cook, R. G., Gorovsky, M. A. and Allis, C. D.,** Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *Proc Natl Acad Sci U S A* 1998. **95**: 7480-7484.

62 **Gurley, L. R., D'Anna, J. A., Barham, S. S., Deaven, L. L. and Tobey, R. A.,** Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur J Biochem* 1978. **84**: 1-15.

63 **Goto, H., Tomono, Y., Ajiro, K., Kosako, H., Fujita, M., Sakurai, M., Okawa, K., Iwamatsu, A., Okigaki, T., Takahashi, T. and Inagaki, M.,** Identification of a novel phosphorylation site on histone H3 coupled with mitotic chromosome condensation. *J Biol Chem* 1999. **274**: 25543-25549.

64 **Preuss, U., Landsberg, G. and Scheidtmann, K. H.,** Novel mitosis-specific phosphorylation of histone H3 at Thr11 mediated by Dlk/ZIP kinase. *Nucleic Acids Res* 2003. **31**: 878-885.

65 **Hendzel, M. J., Wei, Y., Mancini, M. A., Van Hooser, A., Ranalli, T., Brinkley, B. R., Bazett-Jones, D. P. and Allis, C. D.,** Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* 1997. **106**: 348-360.

66 **Van Hooser, A., Goodrich, D. W., Allis, C. D., Brinkley, B. R. and Mancini, M. A.,** Histone H3 phosphorylation is required for the initiation, but not

maintenance, of mammalian chromosome condensation. *J Cell Sci* 1998. **111 (Pt 23)**: 3497-3506.

67 **Hsu, J. Y., Sun, Z. W., Li, X., Reuben, M., Tatchell, K., Bishop, D. K., Grushcow, J. M., Brame, C. J., Caldwell, J. A., Hunt, D. F., Lin, R., Smith, M. M. and Allis, C. D.,** Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* 2000. **102**: 279-291.

68 **Mahadevan, L. C., Willis, A. C. and Barratt, M. J.,** Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. *Cell* 1991. **65**: 775-783.

69 **Edmondson, D. G., Davie, J. K., Zhou, J., Mirnikjoo, B., Tatchell, K. and Dent, S. Y.,** Site-specific loss of acetylation upon phosphorylation of histone H3. *J Biol Chem* 2002. **277**: 29496-29502.

70 **Hill, C. S., Rimmer, J. M., Green, B. N., Finch, J. T. and Thomas, J. O.,** Histone-DNA interactions and their modulation by phosphorylation of -Ser-Pro-X-Lys/Arg- motifs. *Embo J* 1991. **10**: 1939-1948.

71 **Miller, J. and Gordon, C.,** The regulation of proteasome degradation by multi-ubiquitin chain binding proteins. *FEBS Lett* 2005. **579**: 3224-3230.

72 **Osley, M. A.,** H2B ubiquitylation: the end is in sight. *Biochim Biophys Acta* 2004. **1677**: 74-78.

73 **Zhang, Y.,** Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* 2003. **17**: 2733-2740.

74 **Nickel, B. E., Allis, C. D. and Davie, J. R.,** Ubiquitinated histone H2B is preferentially located in transcriptionally active chromatin. *Biochemistry* 1989. **28**: 958-963.

75 **Levinger, L. and Varshavsky, A.,** Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. *Cell* 1982. **28**: 375-385.

76 **Hensold, J. O., Swerdlow, P. S. and Housman, D. E.,** A transient increase in histone H2A ubiquitination is coincident with the onset of erythroleukemic cell differentiation. *Blood* 1988. **71**: 1153-1156.

77 **Davie, J. R., Lin, R. and Allis, C. D.,** Timing of the appearance of ubiquitinated histones in developing new macronuclei of Tetrahymena thermophila. *Biochem Cell Biol* 1991. **69**: 66-71.

78 **Davie, J. R. and Murphy, L. C.,** Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry* 1990. **29**: 4752-4757.

79 **Barsoum, J. and Varshavsky, A.,** Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. *J Biol Chem* 1985. **260**: 7688-7697.

80 **Henry, K. W., Wyce, A., Lo, W. S., Duggan, L. J., Emre, N. C., Kao, C. F., Pillus, L., Shilatifard, A., Osley, M. A. and Berger, S. L.,** Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* 2003. **17**: 2648-2663.

81 **Sun, Z. W. and Allis, C. D.,** Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 2002. **418**: 104-108.

82 **Ng, H. H., Xu, R. M., Zhang, Y. and Struhl, K.,** Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. *J Biol Chem* 2002. **277**: 34655-34657.

83 **Turner, S. D., Ricci, A. R., Petropoulos, H., Genereaux, J., Skerjanc, I. S. and Brandl, C. J.,** The E2 ubiquitin conjugase Rad6 is required for the ArgR/Mcm1 repression of ARG1 transcription. *Mol Cell Biol* 2002. **22**: 4011-4019.

84 **Kleinschmidt, A. M. and Martinson, H. G.,** Structure of nucleosome core particles containing uH2A (A24). *Nucleic Acids Res* 1981. **9**: 2423-2431.

85 **Ballal, N. R., Kang, Y. J., Olson, M. O. and Busch, H.,** Changes in nucleolar proteins and their phosphorylation patterns during liver regeneration. *J Biol Chem* 1975. **250**: 5921-5925.

86 **Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y.,** Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 2004. **431**: 873-878.

87 **Jason, L. J., Finn, R. M., Lindsey, G. and Ausio, J.,** Histone H2A ubiquitination does not preclude histone H1 binding, but it facilitates its association with the nucleosome. *J Biol Chem* 2005. **280**: 4975-4982.

88 **Kingston, R. E. and Narlikar, G. J.,** ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* 1999. **13**: 2339-2352.

89 **Flaus, A. and Owen-Hughes, T.,** Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? *Curr Opin Genet Dev* 2004. **14**: 165-173.

90 **Neigeborn, L. and Carlson, M.,** Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics* 1984. **108**: 845-858.

91 **Stern, M., Jensen, R. and Herskowitz, I.,** Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol* 1984. **178**: 853-868.

92 **Hirschhorn, J. N., Brown, S. A., Clark, C. D. and Winston, F.,** Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev* 1992. **6**: 2288-2298.

93 **Sudarsanam, P., Iyer, V. R., Brown, P. O. and Winston, F.,** Whole-genome expression analysis of snf/swi mutants of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 2000. **97**: 3364-3369.

94 **Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P. and Reinberg, D.,** Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). *Mol Cell Biol* 2002. **22**: 835-848.

95 **Sif, S., Saurin, A. J., Imbalzano, A. N. and Kingston, R. E.,** Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev* 2001. **15**: 603-618.

96 **Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B. and Kornberg, R. D.,** RSC, an essential, abundant chromatin-remodeling complex. *Cell* 1996. **87**: 1249-1260.

97 **Angus-Hill, M. L., Schlichter, A., Roberts, D., Erdjument-Bromage, H., Tempst, P. and Cairns, B. R.,** A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. *Mol Cell* 2001. **7**: 741-751.

98 **Moreira, J. M. and Holmberg, S.,** Transcriptional repression of the yeast CHA1 gene requires the chromatin-remodeling complex RSC. *Embo J* 1999. **18**: 2836-2844.

99 **Reyes, J. C., Barra, J., Muchardt, C., Camus, A., Babinet, C. and Yaniv, M.,** Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). *Embo J* 1998. **17**: 6979-6991.

100 **Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G. and Magnuson, T.,** A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* 2000. **6**: 1287-1295.

101 **Bochar, D. A., Wang, L., Beniya, H., Kinev, A., Xue, Y., Lane, W. S., Wang, W., Kashanchi, F. and Shiekhattar, R.,** BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* 2000. **102**: 257-265.

102 **Boyer, L. A., Langer, M. R., Crowley, K. A., Tan, S., Denu, J. M. and Peterson, C. L.,** Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell* 2002. **10**: 935-942.

103 **de la Cruz, X., Lois, S., Sanchez-Molina, S. and Martinez-Balbas, M. A.,** Do protein motifs read the histone code? *Bioessays* 2005. **27**: 164-175.

104 **Tsukiyama, T., Daniel, C., Tamkun, J. and Wu, C.,** ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 1995. **83**: 1021-1026.

105 **Tsukiyama, T. and Wu, C.,** Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* 1995. **83**: 1011-1020.

106 **Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R. and Kadonaga, J. T.,** ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* 1997. **90**: 145-155.

107 **Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M. and Becker, P. B.,** Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* 1997. **388**: 598-602.

108 **Tsukiyama, T., Palmer, J., Landel, C. C., Shiloach, J. and Wu, C.,** Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. *Genes Dev* 1999. **13**: 686-697.

109 **Goldmark, J. P., Fazzio, T. G., Estep, P. W., Church, G. M. and Tsukiyama, T.,** The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell* 2000. **103**: 423-433.

110 **Rundlett, S. E., Carmen, A. A., Suka, N., Turner, B. M. and Grunstein, M.,** Transcriptional repression by UME6 involves deacetylation of lysine 5 of histone H4 by RPD3. *Nature* 1998. **392**: 831-835.

111 **Kadosh, D. and Struhl, K.,** Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* 1997. **89**: 365-371.

112 **Kadosh, D. and Struhl, K.,** Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo. *Mol Cell Biol* 1998. **18**: 5121-5127.

113 **Fazzio, T. G., Kooperberg, C., Goldmark, J. P., Neal, C., Basom, R., Delrow, J. and Tsukiyama, T.,** Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol Cell Biol* 2001. **21**: 6450-6460.

114 **Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N. and Mellor, J.,** Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell* 2003. **115**: 425-435.

115 **Lazzaro, M. A. and Picketts, D. J.,** Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. *J Neurochem* 2001. **77**: 1145-1156.
116 **Stopka, T. and Skoultchi, A. I.,** The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci U S A* 2003. **100**: 14097-14102.
117 **Barak, O., Lazzaro, M. A., Lane, W. S., Speicher, D. W., Picketts, D. J. and Shiekhattar, R.,** Isolation of human NURF: a regulator of Engrailed gene expression. *Embo J* 2003. **22**: 6089-6100.
118 **Banting, G. S., Barak, O., Ames, T. M., Burnham, A. C., Kardel, M. D., Cooch, N. S., Davidson, C. E., Godbout, R., McDermid, H. E. and Shiekhattar, R.,** CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum Mol Genet* 2005. **14**: 513-524.
119 **Stokes, D. G., Tartof, K. D. and Perry, R. P.,** CHD1 is concentrated in interbands and puffed regions of Drosophila polytene chromosomes. *Proc Natl Acad Sci U S A* 1996. **93**: 7137-7142.
120 **Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd and Grant, P. A.,** Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 2005. **433**: 434-438.
121 **Lusser, A., Urwin, D. L. and Kadonaga, J. T.,** Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol* 2005. **12**: 160-166.
122 **Feng, Q. and Zhang, Y.,** The NuRD complex: linking histone modification to nucleosome remodeling. *Curr Top Microbiol Immunol* 2003. **274**: 269-290.
123 **Marhold, J., Kramer, K., Kremmer, E. and Lyko, F.,** The Drosophila MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA. *Development* 2004. **131**: 6033-6039.
124 **Williams, C. J., Naito, T., Arco, P. G., Seavitt, J. R., Cashman, S. M., De Souza, B., Qi, X., Keables, P., Von Andrian, U. H. and Georgopoulos, K.,** The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development. *Immunity* 2004. **20**: 719-733.
125 **Tran, H. G., Steger, D. J., Iyer, V. R. and Johnson, A. D.,** The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *Embo J* 2000. **19**: 2323-2331.
126 **Simic, R., Lindstrom, D. L., Tran, H. G., Roinick, K. L., Costa, P. J., Johnson, A. D., Hartzog, G. A. and Arndt, K. M.,** Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *Embo J* 2003. **22**: 1846-1856.
127 **Ebbert, R., Birkmann, A. and Schuller, H. J.,** The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol Microbiol* 1999. **32**: 741-751.
128 **Hagerman, P. J.,** Flexibility of DNA. *Annu Rev Biophys Biophys Chem* 1988. **17**: 265-286.
129 **Richmond, T. J. and Davey, C. A.,** The structure of DNA in the nucleosome core. *Nature* 2003. **423**: 145-150.
130 **Shore, D., Langowski, J. and Baldwin, R. L.,** DNA flexibility studied by covalent closure of short fragments into circles. *Proc Natl Acad Sci U S A* 1981. **78**: 4833-4837.

131 **Widom, J.,** Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys* 2001. **34**: 269-324.
132 **Luger, K. and Richmond, T. J.,** DNA binding within the nucleosome core. *Curr Opin Struct Biol* 1998. **8**: 33-40.
133 **Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. and Richmond, T. J.,** Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* 2002. **319**: 1097-1113.
134 **Langst, G. and Becker, P. B.,** Nucleosome remodeling: one mechanism, many phenomena? *Biochim Biophys Acta* 2004. **1677**: 58-63.
135 **Widom, J.,** Structure, dynamics, and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct* 1998. **27**: 285-327.
136 **Polach, K. J. and Widom, J.,** Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J Mol Biol* 1995. **254**: 130-149.
137 **Polach, K. J. and Widom, J.,** A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J Mol Biol* 1996. **258**: 800-812.
138 **Taylor, I. C., Workman, J. L., Schuetz, T. J. and Kingston, R. E.,** Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. *Genes Dev* 1991. **5**: 1285-1298.
139 **Owen-Hughes, T. and Workman, J. L.,** Remodeling the chromatin structure of a nucleosome array by transcription factor-targeted trans-displacement of histones. *Embo J* 1996. **15**: 4702-4712.
140 **Ura, K., Hayes, J. J. and Wolffe, A. P.,** A positive role for nucleosome mobility in the transcriptional activity of chromatin templates: restriction by linker histones. *Embo J* 1995. **14**: 3752-3765.
141 **Anderson, J. D., Thastrom, A. and Widom, J.,** Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation. *Mol Cell Biol* 2002. **22**: 7147-7157.
142 **Li, G. and Widom, J.,** Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol* 2004. **11**: 763-769.
143 **Saha, A., Wittmeyer, J. and Cairns, B. R.,** Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev* 2002. **16**: 2120-2134.
144 **Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D. and Owen-Hughes, T.,** Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol Cell Biol* 2003. **23**: 1935-1945.
145 **Fyodorov, D. V. and Kadonaga, J. T.,** Dynamics of ATP-dependent chromatin assembly by ACF. *Nature* 2002. **418**: 897-900.
146 **Langst, G. and Becker, P. B.,** ISWI induces nucleosome sliding on nicked DNA. *Mol Cell* 2001. **8**: 1085-1092.
147 **Aoyagi, S. and Hayes, J. J.,** hSWI/SNF-catalyzed nucleosome sliding does not occur solely via a twist-diffusion mechanism. *Mol Cell Biol* 2002. **22**: 7484-7490.
148 **Aoyagi, S., Wade, P. A. and Hayes, J. J.,** Nucleosome sliding induced by the xMi-2 complex does not occur exclusively via a simple twist-diffusion mechanism. *J Biol Chem* 2003. **278**: 30562-30568.
149 **Lorch, Y., Davis, B. and Kornberg, R. D.,** Chromatin remodeling by DNA bending, not twisting. *Proc Natl Acad Sci U S A* 2005. **102**: 1329-1332.

150 **Boeger, H., Griesenbeck, J., Strattan, J. S. and Kornberg, R. D.,** Nucleosomes unfold completely at a transcriptionally active promoter. *Mol Cell* 2003. **11**: 1587-1598.
151 **Reinke, H. and Horz, W.,** Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell* 2003. **11**: 1599-1607.
152 **Kassabov, S. R., Zhang, B., Persinger, J. and Bartholomew, B.,** SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell* 2003. **11**: 391-403.
153 **Bazett-Jones, D. P., Cote, J., Landel, C. C., Peterson, C. L. and Workman, J. L.,** The SWI/SNF complex creates loop domains in DNA and polynucleosome arrays and can disrupt DNA-histone contacts within these domains. *Mol Cell Biol* 1999. **19**: 1470-1478.
154 **Boyer, L. A., Shao, X., Ebright, R. H. and Peterson, C. L.,** Roles of the histone H2A-H2B dimers and the (H3-H4)(2) tetramer in nucleosome remodeling by the SWI-SNF complex. *J Biol Chem* 2000. **275**: 11545-11552.
155 **Kobor, M. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., Madhani, H. D. and Rine, J.,** A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol* 2004. **2**: E131.
156 **Meneghini, M. D., Wu, M. and Madhani, H. D.,** Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* 2003. **112**: 725-736.
157 **Orphanides, G. and Reinberg, D.,** RNA polymerase II elongation through chromatin. *Nature* 2000. **407**: 471-475.
158 **Kireeva, M. L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M. and Studitsky, V. M.,** Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell* 2002. **9**: 541-552.
159 **Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S. and Reinberg, D.,** FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* 1998. **92**: 105-116.
160 **Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M. and Reinberg, D.,** FACT facilitates transcription-dependent nucleosome alteration. *Science* 2003. **301**: 1090-1093.
161 **Schwabish, M. A. and Struhl, K.,** Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol Cell Biol* 2004. **24**: 10111-10117.
162 **Baer, B. W. and Rhodes, D.,** Eukaryotic RNA polymerase II binds to nucleosome cores from transcribed genes. *Nature* 1983. **301**: 482-488.
163 **Hayes, J. J. and Wolffe, A. P.,** Histones H2A/H2B inhibit the interaction of transcription factor IIIA with the Xenopus borealis somatic 5S RNA gene in a nucleosome. *Proc Natl Acad Sci U S A* 1992. **89**: 1229-1233.
164 **Morris, S. A., Shibata, Y., Noma, K., Tsukamoto, Y., Warren, E., Temple, B., Grewal, S. I. and Strahl, B. D.,** Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. *Eukaryot Cell* 2005. **4**: 1446-1454.
165 **Cho, H., Orphanides, G., Sun, X., Yang, X. J., Ogryzko, V., Lees, E., Nakatani, Y. and Reinberg, D.,** A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol Cell Biol* 1998. **18**: 5355-5363.
166 **Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F. and**

Shilatifard, A., The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* 2003. **11**: 721-729.

167 **Kamakaka, R. T. and Biggins, S.,** Histone variants: deviants? *Genes Dev* 2005. **19**: 295-310.

168 **Rangwala, S. H. and Richards, E. J.,** The value-added genome: building and maintaining genomic cytosine methylation landscapes. *Curr Opin Genet Dev* 2004. **14**: 686-691.

169 **Lippman, Z. and Martienssen, R.,** The role of RNA interference in heterochromatic silencing. *Nature* 2004. **431**: 364-370.

170 **Harikrishnan, K. N., Chow, M. Z., Baker, E. K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J. M., Jones, P. L., Sif, S. and El-Osta, A.,** Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* 2005. **37**: 254-264.

171 **Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. and Peters, A. H.,** Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 2003. **13**: 1192-1200.

172 **Fuks, F., Hurd, P. J., Deplus, R. and Kouzarides, T.,** The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 2003. **31**: 2305-2312.

173 **Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P. and Kouzarides, T.,** The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* 2003. **278**: 4035-4040.

174 **Nguyen, C. T., Weisenberger, D. J., Velicescu, M., Gonzales, F. A., Lin, J. C., Liang, G. and Jones, P. A.,** Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine. *Cancer Res* 2002. **62**: 6456-6461.

175 **Jeddeloh, J. A., Stokes, T. L. and Richards, E. J.,** Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 1999. **22**: 94-97.

176 **Johnson, L., Cao, X. and Jacobsen, S.,** Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 2002. **12**: 1360-1367.

177 **Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. and Martienssen, R. A.,** Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science* 2002. **297**: 1871-1873.

178 **Mutskov, V. and Felsenfeld, G.,** Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *Embo J* 2004. **23**: 138-149.

179 **Mermoud, J. E., Popova, B., Peters, A. H., Jenuwein, T. and Brockdorff, N.,** Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. *Curr Biol* 2002. **12**: 247-251.

180 **Hermann, A., Goyal, R. and Jeltsch, A.,** The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 2004. **279**: 48350-48359.

181 **Pradhan, S., Bacolla, A., Wells, R. D. and Roberts, R. J.,** Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* 1999. **274**: 33002-33010.

182 **Belmont, A. S.,** Mitotic chromosome scaffold structure: new approaches to an old controversy. *Proc Natl Acad Sci U S A* 2002. **99**: 15855-15857.
183 **Paulson, J. R. and Laemmli, U. K.,** The structure of histone-depleted metaphase chromosomes. *Cell* 1977. **12**: 817-828.
184 **Laemmli, U. K., Cheng, S. M., Adolph, K. W., Paulson, J. R., Brown, J. A. and Baumbach, W. R.,** Metaphase chromosome structure: the role of nonhistone proteins. *Cold Spring Harb Symp Quant Biol* 1978. **42 Pt 1**: 351-360.
185 **Earnshaw, W. C. and Heck, M. M.,** Localization of topoisomerase II in mitotic chromosomes. *J Cell Biol* 1985. **100**: 1716-1725.
186 **Gasser, S. M., Laroche, T., Falquet, J., Boy de la Tour, E. and Laemmli, U. K.,** Metaphase chromosome structure. Involvement of topoisomerase II. *J Mol Biol* 1986. **188**: 613-629.
187 **Saitoh, N., Goldberg, I. G., Wood, E. R. and Earnshaw, W. C.,** ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. *J Cell Biol* 1994. **127**: 303-318.
188 **Strunnikov, A. V., Hogan, E. and Koshland, D.,** SMC2, a Saccharomyces cerevisiae gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes Dev* 1995. **9**: 587-599.
189 **Razin, S. V.,** The nuclear matrix and chromosomal DNA loops: is their any correlation between partitioning of the genome into loops and functional domains? *Cell Mol Biol Lett* 2001. **6**: 59-69.
190 **Christensen, M. O., Larsen, M. K., Barthelmes, H. U., Hock, R., Andersen, C. L., Kjeldsen, E., Knudsen, B. R., Westergaard, O., Boege, F. and Mielke, C.,** Dynamics of human DNA topoisomerases IIalpha and IIbeta in living cells. *J Cell Biol* 2002. **157**: 31-44.
191 **Poirier, M. G. and Marko, J. F.,** Mitotic chromosomes are chromatin networks without a mechanically contiguous protein scaffold. *Proc Natl Acad Sci U S A* 2002. **99**: 15393-15397.
192 **Berezney, R. and Coffey, D. S.,** Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei. *J Cell Biol* 1977. **73**: 616-637.
193 **Mika, S. and Rost, B.,** NMPdb: Database of Nuclear Matrix Proteins. *Nucleic Acids Res* 2005. **33**: D160-163.
194 **Tan, J. H., Wooley, J. C. and LeStourgeon, W. M.,** Nuclear matrix-like filaments and fibrogranular complexes form through the rearrangement of specific nuclear ribonucleoproteins. *Mol Biol Cell* 2000. **11**: 1547-1554.
195 **Pederson, T.,** Thinking about a nuclear matrix. *J Mol Biol* 1998. **277**: 147-159.
196 **Pederson, T.,** Half a century of "the nuclear matrix". *Mol Biol Cell* 2000. **11**: 799-805.
197 **Barboro, P., D'Arrigo, C., Mormino, M., Coradeghini, R., Parodi, S., Patrone, E. and Balbi, C.,** An intranuclear frame for chromatin compartmentalization and higher-order folding. *J Cell Biochem* 2003. **88**: 113-120.
198 **Fey, E. G., Ornelles, D. A. and Penman, S.,** Association of RNA with the cytoskeleton and the nuclear matrix. *J Cell Sci Suppl* 1986. **5**: 99-119.
199 **van Venrooij, W. J. and Janssen, D. B.,** HnRNP particles. *Mol Biol Rep* 1978. **4**: 3-8.

200 **Berezney, R.,** Fractionation of the nuclear matrix. I. Partial separation into matrix protein fibrils and a residual ribonucleoprotein fraction. *J Cell Biol* 1980. **85**: 641-650.

201 **Heaphy, S., Finch, J. T., Gait, M. J., Karn, J. and Singh, M.,** Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc Natl Acad Sci U S A* 1991. **88**: 7366-7370.

202 **Shumaker, D. K., Kuczmarski, E. R. and Goldman, R. D.,** The nucleoskeleton: lamins and actin are major players in essential nuclear functions. *Curr Opin Cell Biol* 2003. **15**: 358-366.

203 **Nalepa, G. and Harper, J. W.,** Visualization of a highly organized intranuclear network of filaments in living mammalian cells. *Cell Motil Cytoskeleton* 2004. **59**: 94-108.

204 **Hancock, R.,** Internal organisation of the nucleus: assembly of compartments by macromolecular crowding and the nuclear matrix model. *Biol Cell* 2004. **96**: 595-601.

205 **Nickerson, J.,** Experimental observations of a nuclear matrix. *J Cell Sci* 2001. **114**: 463-474.

206 **Saitoh, Y. and Laemmli, U. K.,** Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* 1994. **76**: 609-622.

207 **Mirkovitch, J., Gasser, S. M. and Laemmli, U. K.,** Scaffold attachment of DNA loops in metaphase chromosomes. *J Mol Biol* 1988. **200**: 101-109.

208 **Razin, S. V.,** Functional architecture of chromosomal DNA domains. *Crit Rev Eukaryot Gene Expr* 1996. **6**: 247-269.

209 **Cockerill, P. N. and Garrard, W. T.,** Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* 1986. **44**: 273-282.

210 **Namciu, S. J. and Fournier, R. E.,** Human matrix attachment regions are necessary for the establishment but not the maintenance of transgene insulation in Drosophila melanogaster. *Mol Cell Biol* 2004. **24**: 10236-10245.

211 **Stief, A., Winter, D. M., Stratling, W. H. and Sippel, A. E.,** A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* 1989. **341**: 343-345.

212 **Kalos, M. and Fournier, R. E.,** Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol Cell Biol* 1995. **15**: 198-207.

213 **Kellum, R. and Schedl, P.,** A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 1991. **64**: 941-950.

214 **Forrester, W. C., Fernandez, L. A. and Grosschedl, R.,** Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev* 1999. **13**: 3003-3014.

215 **Mesner, L. D., Hamlin, J. L. and Dijkwel, P. A.,** The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation. *Proc Natl Acad Sci U S A* 2003. **100**: 3281-3286.

216 **Dijkwel, P. A., Wenink, P. W. and Poddighe, J.,** Permanent attachment of replication origins to the nuclear matrix in BHK-cells. *Nucleic Acids Res* 1986. **14**: 3241-3249.

217 **Carri, M. T., Micheli, G., Graziano, E., Pace, T. and Buongiorno-Nardelli, M.,** The relationship between chromosomal origins of replication and the nuclear matrix during the cell cycle. *Exp Cell Res* 1986. **164**: 426-436.

218 **Glazko, G. V., Koonin, E. V., Rogozin, I. B. and Shabalina, S. A.,** A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet* 2003. **19**: 119-124.

219 **Weber, M., Hagege, H., Murrell, A., Brunel, C., Reik, W., Cathala, G. and Forne, T.,** Genomic imprinting controls matrix attachment regions in the Igf2 gene. *Mol Cell Biol* 2003. **23**: 8953-8959.

220 **Izaurralde, E., Mirkovitch, J. and Laemmli, U. K.,** Interaction of DNA with nuclear scaffolds in vitro. *J Mol Biol* 1988. **200**: 111-125.

221 **Oh, S. J., Jeong, J. S., Kim, E. H., Yi, N. R., Yi, S. I., Jang, I. C., Kim, Y. S., Suh, S. C., Nahm, B. H. and Kim, J. K.,** Matrix attachment region from the chicken lysozyme locus reduces variability in transgene expression and confers copy number-dependence in transgenic rice plants. *Plant Cell Rep* 2005. **24**: 145-154.

222 **Dickinson, L. A. and Kohwi-Shigematsu, T.,** Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Mol Cell Biol* 1995. **15**: 456-465.

223 **van Drunen, C. M., Sewalt, R. G., Oosterling, R. W., Weisbeek, P. J., Smeekens, S. C. and van Driel, R.,** A bipartite sequence element associated with matrix/scaffold attachment regions. *Nucleic Acids Res* 1999. **27**: 2924-2930.

224 **von Kries, J. P., Buck, F. and Stratling, W. H.,** Chicken MAR binding protein p120 is identical to human heterogeneous nuclear ribonucleoprotein (hnRNP) U. *Nucleic Acids Res* 1994. **22**: 1215-1220.

225 **Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R. and Tucker, P. W.,** The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev* 1995. **9**: 3067-3082.

226 **Dickinson, L. A., Joh, T., Kohwi, Y. and Kohwi-Shigematsu, T.,** A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* 1992. **70**: 631-645.

227 **Adachi, Y., Kas, E. and Laemmli, U. K.,** Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *Embo J* 1989. **8**: 3997-4006.

228 **Luderus, M. E., de Graaf, A., Mattia, E., den Blaauwen, J. L., Grande, M. A., de Jong, L. and van Driel, R.,** Binding of matrix attachment regions to lamin B1. *Cell* 1992. **70**: 949-959.

229 **Romig, H., Fackelmayer, F. O., Renz, A., Ramsperger, U. and Richter, A.,** Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *Embo J* 1992. **11**: 3431-3440.

230 **de Belle, I., Cai, S. and Kohwi-Shigematsu, T.,** The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J Cell Biol* 1998. **141**: 335-348.

231 **Alvarez, J. D., Yasui, D. H., Niida, H., Joh, T., Loh, D. Y. and Kohwi-Shigematsu, T.,** The MAR-binding protein SATB1 orchestrates temporal and

spatial expression of multiple genes during T-cell development. *Genes Dev* 2000. **14**: 521-535.

232 **Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P. and Kohwi-Shigematsu, T.,** SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 2002. **419**: 641-645.

233 **Cai, S., Han, H. J. and Kohwi-Shigematsu, T.,** Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* 2003. **34**: 42-51.

234 **Rabl, C.** *Morphol, Jahrb* 1885. **10**: 214-330.

235 **Boveri, T.** *Arch. Exp. Zellforsch.* 1909. **3**: 181-268.

236 **Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V. and Zorn, C.,** Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet* 1982. **60**: 46-56.

237 **Cremer, T., Kupper, K., Dietzel, S. and Fakan, S.,** Higher order chromatin architecture in the cell nucleus: on the way from structure to function. *Biol Cell* 2004. **96**: 555-567.

238 **Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T. and Lichter, P.,** Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res* 1993. **1**: 93-106.

239 **Misteli, T.,** Protein dynamics: implications for nuclear architecture and gene expression. *Science* 2001. **291**: 843-847.

240 **Abranches, R., Beven, A. F., Aragon-Alcaide, L. and Shaw, P. J.,** Transcription sites are not correlated with chromosome territories in wheat nuclei. *J Cell Biol* 1998. **143**: 5-12.

241 **Verschure, P. J., van Der Kraan, I., Manders, E. M. and van Driel, R.,** Spatial relationship between transcription sites and chromosome territories. *J Cell Biol* 1999. **147**: 13-24.

242 **Visser, A. E., Jaunin, F., Fakan, S. and Aten, J. A.,** High resolution analysis of interphase chromosome domains. *J Cell Sci* 2000. **113 (Pt 14)**: 2585-2593.

243 **Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., Goldsworthy, M., Powis, S. H., Ragoussis, J., Trowsdale, J. and Sheer, D.,** Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* 2000. **113 (Pt 9)**: 1565-1576.

244 **Williams, R. R., Broad, S., Sheer, D. and Ragoussis, J.,** Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res* 2002. **272**: 163-175.

245 **Lamond, A. I. and Sleeman, J. E.,** Nuclear substructure and dynamics. *Curr Biol* 2003. **13**: R825-828.

246 **Lorkovic, Z. J. and Barta, A.,** Compartmentalization of the splicing machinery in plant cell nuclei. *Trends Plant Sci* 2004. **9**: 565-568.

247 **Misteli, T., Caceres, J. F. and Spector, D. L.,** The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 1997. **387**: 523-527.

248 **Eils, R., Gerlich, D., Tvarusko, W., Spector, D. L. and Misteli, T.,** Quantitative imaging of pre-mRNA splicing factors in living cells. *Mol Biol Cell* 2000. **11**: 413-418.

249 **Sleeman, J. E. and Lamond, A. I.,** Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol* 1999. **9**: 1065-1074.

250 **Narayanan, A., Speckmann, W., Terns, R. and Terns, M. P.,** Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell* 1999. **10**: 2131-2147.

251 **Parada, L. and Misteli, T.,** Chromosome positioning in the interphase nucleus. *Trends Cell Biol* 2002. **12**: 425-432.

252 **Chubb, J. R., Boyle, S., Perry, P. and Bickmore, W. A.,** Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* 2002. **12**: 439-445.

253 **Polioudaki, H., Kourmouli, N., Drosou, V., Bakou, A., Theodoropoulos, P. A., Singh, P. B., Giannakouros, T. and Georgatos, S. D.,** Histones H3/H4 form a tight complex with the inner nuclear membrane protein LBR and heterochromatin protein 1. *EMBO Rep* 2001. **2**: 920-925.

254 **Mattout-Drubezki, A. and Gruenbaum, Y.,** Dynamic interactions of nuclear lamina proteins with chromatin and transcriptional machinery. *Cell Mol Life Sci* 2003. **60**: 2053-2063.

255 **Furukawa, K., Glass, C. and Kondo, T.,** Characterization of the chromatin binding activity of lamina-associated polypeptide (LAP) 2. *Biochem Biophys Res Commun* 1997. **238**: 240-246.

256 **Neri, L. M., Raymond, Y., Giordano, A., Borgatti, P., Marchisio, M., Capitani, S. and Martelli, A. M.,** Spatial distribution of lamin A and B1 in the K562 cell nuclear matrix stabilized with metal ions. *J Cell Biochem* 1999. **75**: 36-45.

257 **Marshall, W. F., Straight, A., Marko, J. F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A. W., Agard, D. A. and Sedat, J. W.,** Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr Biol* 1997. **7**: 930-939.

258 **Iborra, F. J., Pombo, A., Jackson, D. A. and Cook, P. R.,** Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci* 1996. **109 (Pt 6)**: 1427-1436.

259 **Jackson, D. A., Iborra, F. J., Manders, E. M. and Cook, P. R.,** Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* 1998. **9**: 1523-1536.

260 **Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W. and Fraser, P.,** Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 2004. **36**: 1065-1071.

261 **Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. and Flavell, R. A.,** Interchromosomal associations between alternatively expressed loci. *Nature* 2005. **435**: 637-645.

262 **Henikoff, S., Jackson, J. M. and Talbert, P. B.,** Distance and pairing effects on the brownDominant heterochromatic element in Drosophila. *Genetics* 1995. **140**: 1007-1017.

263 **Csink, A. K. and Henikoff, S.,** Genetic modification of heterochromatic association and nuclear organization in Drosophila. *Nature* 1996. **381**: 529-531.

264 **Corcoran, A. E.,** Immunoglobulin locus silencing and allelic exclusion. *Semin Immunol* 2005. **17**: 141-154.

265 **Skok, J. A., Brown, K. E., Azuara, V., Caparros, M. L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R. P., Merkenschlager, M. and Fisher, A. G.,** Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat Immunol* 2001. **2**: 848-854.

266 **Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M. and Fisher, A. G.,** Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell* 1999. **3**: 207-217.

267 **Brown, K. E., Guest, S. S., Smale, S. T., Hahm, K., Merkenschlager, M. and Fisher, A. G.,** Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 1997. **91**: 845-854.

268 **Lundgren, M., Chow, C. M., Sabbattini, P., Georgiou, A., Minaee, S. and Dillon, N.,** Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 2000. **103**: 733-743.

269 **Brown, K. E., Amoils, S., Horn, J. M., Buckle, V. J., Higgs, D. R., Merkenschlager, M. and Fisher, A. G.,** Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells. *Nat Cell Biol* 2001. **3**: 602-606.

270 **Heitz, E.,** Das heterochromatin der moose. *I. Jahrb. Wiss. Botanik.* 1928. **69**: 762-818.

271 **Kellum, R.,** Is HP1 an RNA detector that functions both in repression and activation? *J Cell Biol* 2003. **161**: 671-672.

272 **Bailis, J. M. and Forsburg, S. L.,** It's all in the timing: linking S phase to chromatin structure and chromosome dynamics. *Cell Cycle* 2003. **2**: 303-306.

273 **Gilbert, D. M.,** Replication timing and transcriptional control: beyond cause and effect. *Curr Opin Cell Biol* 2002. **14**: 377-383.

274 **Wintersberger, E.,** Why is there late replication? *Chromosoma* 2000. **109**: 300-307.

275 **Zhimulev, I. F., Belyaeva, E. S., Bolshakov, V. N. and Mal'ceva, N. I.,** Position-effect variegation and intercalary heterochromatin: a comparative study. *Chromosoma* 1989. **98**: 378-387.

276 **Rountree, M. R., Bachman, K. E. and Baylin, S. B.,** DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet* 2000. **25**: 269-277.

277 **Dillon, N.,** Heterochromatin structure and function. *Biol Cell* 2004. **96**: 631-637.

278 **Maison, C. and Almouzni, G.,** HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* 2004. **5**: 296-304.

279 **Minc, E., Courvalin, J. C. and Buendia, B.,** HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. *Cytogenet Cell Genet* 2000. **90**: 279-284.

280 **Hayakawa, T., Haraguchi, T., Masumoto, H. and Hiraoka, Y.,** Cell cycle behavior of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase. *J Cell Sci* 2003. **116**: 3327-3338.

281 **Maison, C., Bailly, D., Peters, A. H., Quivy, J. P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T. and Almouzni, G.,** Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* 2002. **30**: 329-334.

282 **Gilbert, N. and Allan, J.,** Distinctive higher-order chromatin structure at mammalian centromeres. *Proc Natl Acad Sci U S A* 2001. **98**: 11949-11954.

283 **Grigoryev, S. A.,** Keeping fingers crossed: heterochromatin spreading through interdigitation of nucleosome arrays. *FEBS Lett* 2004. **564**: 4-8.

284 **Fyodorov, D. V., Blower, M. D., Karpen, G. H. and Kadonaga, J. T.,** Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev* 2004. **18**: 170-183.

285 **Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A. and Hansen, J. C.,** Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J Biol Chem* 2003. **278**: 32181-32188.

286 **Li, Y., Danzer, J. R., Alvarez, P., Belmont, A. S. and Wallrath, L. L.,** Effects of tethering HP1 to euchromatic regions of the Drosophila genome. *Development* 2003. **130**: 1817-1824.

287 **Thiru, A., Nietlispach, D., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V. and Laue, E. D.,** Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *Embo J* 2004. **23**: 489-499.

288 **Lavigne, M., Francis, N. J., King, I. F. and Kingston, R. E.,** Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin. *Mol Cell* 2004. **13**: 415-425.

289 **Springhetti, E. M., Istomina, N. E., Whisstock, J. C., Nikitina, T., Woodcock, C. L. and Grigoryev, S. A.,** Role of the M-loop and reactive center loop domains in the folding and bridging of nucleosome arrays by MENT. *J Biol Chem* 2003. **278**: 43384-43393.

290 **Dimitri, P., Corradini, N., Rossi, F. and Verni, F.,** The paradox of functional heterochromatin. *Bioessays* 2005. **27**: 29-41.

291 **Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M. and Gvozdev, V. A.,** Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* 2001. **11**: 1017-1027.

292 **Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A. and Elgin, S. C.,** Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* 2004. **303**: 669-672.

293 **Soppe, W. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I. and Fransz, P. F.,** DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *Embo J* 2002. **21**: 6549-6559.

294 **Lippman, Z., May, B., Yordan, C., Singer, T. and Martienssen, R.,** Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* 2003. **1**: E67.

295 **Pimpinelli, S. and Goday, C.,** Unusual kinetochores and chromatin diminution in Parascaris. *Trends Genet* 1989. **5**: 310-315.

296 **Dernburg, A. F., Sedat, J. W. and Hawley, R. S.,** Direct evidence of a role for heterochromatin in meiotic chromosome segregation. *Cell* 1996. **86**: 135-146.

297 **Karpen, G. H., Le, M. H. and Le, H.,** Centric heterochromatin and the efficiency of achiasmate disjunction in Drosophila female meiosis. *Science* 1996. **273**: 118-122.

298 **Amor, D. J., Kalitsis, P., Sumer, H. and Choo, K. H.,** Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol* 2004. **14**: 359-368.

299 **Schramke, V. and Allshire, R.,** Those interfering little RNAs! Silencing and eliminating chromatin. *Curr Opin Genet Dev* 2004. **14**: 174-180.
300 **Csink, A. K. and Henikoff, S.,** Something from nothing: the evolution and utility of satellite repeats. *Trends Genet* 1998. **14**: 200-204.
301 **Hoskins, R. A., Smith, C. D., Carlson, J. W., Carvalho, A. B., Halpern, A., Kaminker, J. S., Kennedy, C., Mungall, C. J., Sullivan, B. A., Sutton, G. G., Yasuhara, J. C., Wakimoto, B. T., Myers, E. W., Celniker, S. E., Rubin, G. M. and Karpen, G. H.,** Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome Biol* 2002. **3**: RESEARCH0085.
302 **Corradini, N., Rossi, F., Verni, F. and Dimitri, P.,** FISH analysis of Drosophila melanogaster heterochromatin using BACs and P elements. *Chromosoma* 2003. **112**: 26-37.
303 **Kuhn, R. M., Clarke, L. and Carbon, J.,** Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA sequence repeats. *Proc Natl Acad Sci U S A* 1991. **88**: 1306-1310.
304 **Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P. B., Kim, M., Jones, K. M., Henikoff, S., Buell, C. R. and Jiang, J.,** Sequencing of a rice centromere uncovers active genes. *Nat Genet* 2004. **36**: 138-145.
305 **Wakimoto, B. T. and Hearn, M. G.,** The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. *Genetics* 1990. **125**: 141-154.
306 **Eberl, D. F., Duyf, B. J. and Hilliker, A. J.,** The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of Drosophila melanogaster. *Genetics* 1993. **134**: 277-292.
307 **Heard, E.,** Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* 2004. **16**: 247-255.
308 **McCarrey, J. R. and Dilworth, D. D.,** Expression of Xist in mouse germ cells correlates with X-chromosome inactivation. *Nat Genet* 1992. **2**: 200-203.
309 **Chadwick, B. P. and Willard, H. F.,** Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Hum Mol Genet* 2001. **10**: 1101-1113.
310 **Goto, T. and Monk, M.,** Regulation of X-chromosome inactivation in development in mice and humans. *Microbiol Mol Biol Rev* 1998. **62**: 362-378.
311 **Grewal, S. I. and Rice, J. C.,** Regulation of heterochromatin by histone methylation and small RNAs. *Curr Opin Cell Biol* 2004. **16**: 230-238.
312 **Bender, J.,** Chromatin-based silencing mechanisms. *Curr Opin Plant Biol* 2004. **7**: 521-526.
313 **Martienssen, R. A.,** Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat Genet* 2003. **35**: 213-214.
314 **Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J.,** Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001. **409**: 363-366.
315 **Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. and Hannon, G. J.,** Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 2001. **293**: 1146-1150.
316 **Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. and Martienssen, R. A.,** Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 2002. **297**: 1833-1837.
317 **Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T. and Oshimura, M.,** Dicer is essential for

formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol* 2004. **6**: 784-791.

318 **Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I. and Moazed, D.,** RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 2004. **303**: 672-676.

319 **Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P. and Moazed, D.,** Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* 2004. **119**: 789-802.

320 **Sugiyama, T., Cam, H., Verdel, A., Moazed, D. and Grewal, S. I.,** RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci U S A* 2005. **102**: 152-157.

321 **Schramke, V. and Allshire, R.,** Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* 2003. **301**: 1069-1074.

322 **Elgin, S. C. and Grewal, S. I.,** Heterochromatin: silence is golden. *Curr Biol* 2003. **13**: R895-898.

323 **Vermaak, D., Ahmad, K. and Henikoff, S.,** Maintenance of chromatin states: an open-and-shut case. *Curr Opin Cell Biol* 2003. **15**: 266-274.

324 **Muller, H. J.,** Types of viable variations induced by X-rays in *Drosophila*. *J. Genet.* 1930. **22**: 299-334.

325 **Schultz, J.,** Variegation in *Drosophila* and the inert chromosome regions. *Proc. Natl Acad. Sci. USA* 1936. **22**: 22-33.

326 **Karpen, G. H.,** Position-effect variegation and the new biology of heterochromatin. *Curr Opin Genet Dev* 1994. **4**: 281-291.

327 **Singh, P. B.,** Molecular mechanisms of cellular determination: their relation to chromatin structure and parental imprinting. *J Cell Sci* 1994. **107 (Pt 10)**: 2653-2668.

328 **Dorer, D. R. and Henikoff, S.,** Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* 1994. **77**: 993-1002.

329 **Sabl, J. F. and Henikoff, S.,** Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. *Genetics* 1996. **142**: 447-458.

330 **Martin-Morris, L. E., Csink, A. K., Dorer, D. R., Talbert, P. B. and Henikoff, S.,** Heterochromatic trans-inactivation of Drosophila white transgenes. *Genetics* 1997. **147**: 671-677.

331 **Festenstein, R. and Kioussis, D.,** Locus control regions and epigenetic chromatin modifiers. *Curr Opin Genet Dev* 2000. **10**: 199-203.

332 **Zhimulev, I. F. and Belyaeva, E. S.,** Intercalary heterochromatin and genetic silencing. *Bioessays* 2003. **25**: 1040-1051.

333 **Wallrath, L. L. and Elgin, S. C.,** Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes Dev* 1995. **9**: 1263-1277.

334 **Sun, F. L., Cuaycong, M. H. and Elgin, S. C.,** Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Mol Cell Biol* 2001. **21**: 2867-2879.

335 **Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R. and Reuter, G.,** Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *Embo J* 2002. **21**: 1121-1131.

336 **Dorer, D. R. and Henikoff, S.,** Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* 1997. **147**: 1181-1190.

337 **Reuter, G. and Wolff, I.,** Isolation of dominant suppressor mutations for position-effect variegation in Drosophila melanogaster. *Mol Gen Genet* 1981. **182**: 516-519.

338 **Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. and Reuter, G.,** The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo J* 1994. **13**: 3822-3831.

339 **Eissenberg, J. C., James, T. C., Foster-Hartnett, D. M., Hartnett, T., Ngan, V. and Elgin, S. C.,** Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proc Natl Acad Sci U S A* 1990. **87**: 9923-9927.

340 **Palmiter, R. D. and Brinster, R. L.,** Germ-line transformation of mice. *Annu Rev Genet* 1986. **20**: 465-499.

341 **Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. and Kioussis, D.,** Locus control region function and heterochromatin-induced position effect variegation. *Science* 1996. **271**: 1123-1125.

342 **Festenstein, R., Sharghi-Namini, S., Fox, M., Roderick, K., Tolaini, M., Norton, T., Saveliev, A., Kioussis, D. and Singh, P.,** Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner. *Nat Genet* 1999. **23**: 457-461.

343 **Saunders, W. S., Chue, C., Goebl, M., Craig, C., Clark, R. F., Powers, J. A., Eissenberg, J. C., Elgin, S. C., Rothfield, N. F. and Earnshaw, W. C.,** Molecular cloning of a human homologue of Drosophila heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity. *J Cell Sci* 1993. **104 (Pt 2)**: 573-582.

344 **Singh, P. B., Miller, J. R., Pearce, J., Kothary, R., Burton, R. D., Paro, R., James, T. C. and Gaunt, S. J.,** A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res* 1991. **19**: 789-794.

345 **Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P. B., Reuter, G. and Jenuwein, T.,** Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *Embo J* 1999. **18**: 1923-1938.

346 **Wreggett, K. A., Hill, F., James, P. S., Hutchings, A., Butcher, G. W. and Singh, P. B.,** A mammalian homologue of Drosophila heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenet Cell Genet* 1994. **66**: 99-103.

347 **Horsley, D., Hutchings, A., Butcher, G. W. and Singh, P. B.,** M32, a murine homologue of Drosophila heterochromatin protein 1 (HP1), localises to euchromatin within interphase nuclei and is largely excluded from constitutive heterochromatin. *Cytogenet Cell Genet* 1996. **73**: 308-311.

348 **Grosveld, F., van Assendelft, G. B., Greaves, D. R. and Kollias, G.,** Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 1987. **51**: 975-985.

349 **Li, Q., Harju, S. and Peterson, K. R.,** Locus control regions: coming of age at a decade plus. *Trends Genet* 1999. **15**: 403-408.

350 **Li, Q., Peterson, K. R., Fang, X. and Stamatoyannopoulos, G.,** Locus control regions. *Blood* 2002. **100**: 3077-3086.

351 **Li, X. G., Liu, D. P. and Liang, C. C.,** Beyond the locus control region: new light on beta-globin locus regulation. *Int J Biochem Cell Biol* 2001. **33**: 914-923.

352 **Levings, P. P. and Bungert, J.,** The human beta-globin locus control region. *Eur J Biochem* 2002. **269**: 1589-1599.

353 **Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. and Grosveld, F. G.,** Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* 1983. **306**: 662-666.

354 **Van der Ploeg, L. H., Konings, A., Oort, M., Roos, D., Bernini, L. and Flavell, R. A.,** gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* 1980. **283**: 637-642.

355 **Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M.,** A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes Dev* 1990. **4**: 1637-1649.

356 **Kollias, G., Wrighton, N., Hurst, J. and Grosveld, F.,** Regulated expression of human A gamma-, beta-, and hybrid gamma beta-globin genes in transgenic mice: manipulation of the developmental expression patterns. *Cell* 1986. **46**: 89-94.

357 **Magram, J., Chada, K. and Costantini, F.,** Developmental regulation of a cloned adult beta-globin gene in transgenic mice. *Nature* 1985. **315**: 338-340.

358 **Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. and Palmiter, R. D.,** Erythroid-specific expression of human beta-globin genes in transgenic mice. *Embo J* 1985. **4**: 1715-1723.

359 **Tuan, D., Solomon, W., Li, Q. and London, I. M.,** The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A* 1985. **82**: 6384-6388.

360 **Li, Q., Zhang, M., Duan, Z. and Stamatoyannopoulos, G.,** Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics* 1999. **61**: 183-193.

361 **Alami, R., Greally, J. M., Tanimoto, K., Hwang, S., Feng, Y. Q., Engel, J. D., Fiering, S. and Bouhassira, E. E.,** Beta-globin YAC transgenes exhibit uniform expression levels but position effect variegation in mice. *Hum Mol Genet* 2000. **9**: 631-636.

362 **Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D. R.,** The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell* 1989. **56**: 969-977.

363 **Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S.,** The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial lacZ gene. *Development* 1996. **122**: 3991-3999.

364 **Bulger, M., van Doorninck, J. H., Saitoh, N., Telling, A., Farrell, C., Bender, M. A., Felsenfeld, G., Axel, R. and Groudine, M.,** Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci U S A* 1999. **96**: 5129-5134.

365 **Bender, M. A., Reik, A., Close, J., Telling, A., Epner, E., Fiering, S., Hardison, R. and Groudine, M.,** Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse beta-globin locus control region. *Blood* 1998. **92**: 4394-4403.

366 **Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N. and Miller, W.,** Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 1997. **205**: 73-94.

367 **Farrell, C. M., Grinberg, A., Huang, S. P., Chen, D., Pichel, J. G., Westphal, H. and Felsenfeld, G.,** A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse beta -globin locus. *Proc Natl Acad Sci U S A* 2000. **97**: 14554-14559.

368 **Farrell, C. M., West, A. G. and Felsenfeld, G.,** Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol Cell Biol* 2002. **22**: 3820-3831.

369 **Fleenor, D. E. and Kaufman, R. E.,** Characterization of the DNase I hypersensitive site 3' of the human beta globin gene domain. *Blood* 1993. **81**: 2781-2790.

370 **Francastel, C., Walters, M. C., Groudine, M. and Martin, D. I.,** A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell* 1999. **99**: 259-269.

371 **Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P.,** Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 1996. **87**: 105-114.

372 **Bender, M. A., Mehaffey, M. G., Telling, A., Hug, B., Ley, T. J., Groudine, M. and Fiering, S.,** Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. *Blood* 2000. **95**: 3600-3604.

373 **Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M. A., Fiering, S., Enver, T., Martin, D. I., Kennedy, M., Keller, G. and Groudine, M.,** The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell* 1998. **2**: 447-455.

374 **Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D. I., Enver, T., Ley, T. J. and Groudine, M.,** Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev* 1995. **9**: 2203-2213.

375 **Bender, M. A., Roach, J. N., Halow, J., Close, J., Alami, R., Bouhassira, E. E., Groudine, M. and Fiering, S. N.,** Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood* 2001. **98**: 2022-2027.

376 **Alami, R., Bender, M. A., Feng, Y. Q., Fiering, S. N., Hug, B. A., Ley, T. J., Groudine, M. and Bouhassira, E. E.,** Deletions within the mouse beta-globin locus control region preferentially reduce beta(min) globin gene expression. *Genomics* 2000. **63**: 417-424.

377 **Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M.,** The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol* 1998. **18**: 5992-6000.

378 **Hug, B. A., Wesselschmidt, R. L., Fiering, S., Bender, M. A., Epner, E., Groudine, M. and Ley, T. J.,** Analysis of mice containing a targeted deletion of beta-globin locus control region 5' hypersensitive site 3. *Mol Cell Biol* 1996. **16**: 2906-2912.

379 **Schubeler, D., Francastel, C., Cimbora, D. M., Reik, A., Martin, D. I. and Groudine, M.,** Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev* 2000. **14**: 940-950.

380 **Vieira, K. F., Levings, P. P., Hill, M. A., Crusselle, V. J., Kang, S. H., Engel, J. D. and Bungert, J.,** Recruitment of transcription complexes to the beta-globin gene locus in vivo and in vitro. *J Biol Chem* 2004. **279**: 50350-50357.

381 **Bender, M. A., Bulger, M., Close, J. and Groudine, M.,** Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol Cell* 2000. **5**: 387-393.

382 **Schubeler, D., Groudine, M. and Bender, M. A.,** The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci U S A* 2001. **98**: 11432-11437.

383 **Tang, Y., Liu, D. P. and Liang, C. C.,** Further understanding of the beta-globin locus regulation at the molecular level: looping or linking models? *Genes Cells* 2002. **7**: 889-900.

384 **Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. and Fraser, P.,** Long-range chromatin regulatory interactions in vivo. *Nat Genet* 2002. **32**: 623-626.

385 **Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F. and de Laat, W.,** The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003. **35**: 190-194.

386 **Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W.,** Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 2002. **10**: 1453-1465.

387 **de Laat, W. and Grosveld, F.,** Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* 2003. **11**: 447-459.

388 **Ostermeier, G. C., Liu, Z., Martins, R. P., Bharadwaj, R. R., Ellis, J., Draghici, S. and Krawetz, S. A.,** Nuclear matrix association of the human beta-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Res* 2003. **31**: 3257-3266.

389 **Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N. J.,** Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* 1997. **11**: 2494-2509.

390 **Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. and Fraser, P.,** Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* 2000. **5**: 377-386.

391 **Plant, K. E., Routledge, S. J. and Proudfoot, N. J.,** Intergenic transcription in the human beta-globin gene cluster. *Mol Cell Biol* 2001. **21**: 6507-6514.

392 **Lang, G., Wotton, D., Owen, M. J., Sewell, W. A., Brown, M. H., Mason, D. Y., Crumpton, M. J. and Kioussis, D.,** The structure of the human CD2 gene and its expression in transgenic mice. *Embo J* 1988. **7**: 1675-1682.

393 **Greaves, D. R., Wilson, F. D., Lang, G. and Kioussis, D.,** Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. *Cell* 1989. **56**: 979-986.

394 **Lang, G., Mamalaki, C., Greenberg, D., Yannoutsos, N. and Kioussis, D.,** Deletion analysis of the human CD2 gene locus control region in transgenic mice. *Nucleic Acids Res* 1991. **19**: 5851-5856.

395 **Lake, R. A., Wotton, D. and Owen, M. J.,** A 3' transcriptional enhancer regulates tissue-specific expression of the human CD2 gene. *Embo J* 1990. **9**: 3129-3136.

396 **Zhuma, T., Tyrrell, R., Sekkali, B., Skavdis, G., Saveliev, A., Tolaini, M., Roderick, K., Norton, T., Smerdon, S., Sedgwick, S., Festenstein, R. and Kioussis, D.,** Human HMG box transcription factor HBP1: a role in hCD2 LCR function. *Embo J* 1999. **18**: 6396-6406.

397 **Kioussis, D. and Festenstein, R.,** Chromatin Structure and Lineage Determination. In **Manroe., J. G. and Rothernberg, E. V.** (Eds.) *Molecular Biology Of B-Cell and T-Cell Developmet.* Humana Press, Totowa, New Jersey 1998, pp 127-146.

398 **Southern, E. M.,** Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975. **98**: 503-517.

399 **Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M.,** Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* 1987. **15**: 10159-10177.

400 **Santoso, B., Ortiz, B. D. and Winoto, A.,** Control of organ-specific demethylation by an element of the T-cell receptor-alpha locus control region. *J Biol Chem* 2000. **275**: 1952-1958.

401 **Gu, H., Zou, Y. R. and Rajewsky, K.,** Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* 1993. **73**: 1155-1164.

402 **Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. and Rajewsky, K.,** Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 1994. **265**: 103-106.

403 **Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E. J., Manning, R. W., Yu, S. H., Mulder, K. L. and Westphal, H.,** Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc Natl Acad Sci U S A* 1992. **89**: 6232-6236.

404 **Kilby, N. J., Snaith, M. R. and Murray, J. A.,** Site-specific recombinases: tools for genome engineering. *Trends Genet* 1993. **9**: 413-421.

405 **Sternberg, N. and Hamilton, D.,** Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol* 1981. **150**: 467-486.

406 **Jin, X. L., Guo, H., Mao, C., Atkins, N., Wang, H., Avasthi, P. P., Tu, Y. T. and Li, Y.,** Emx1-specific expression of foreign genes using "knock-in" approach. *Biochem Biophys Res Commun* 2000. **270**: 978-982.

407 **Michael, S. K., Brennan, J. and Robertson, E. J.,** Efficient gene-specific expression of cre recombinase in the mouse embryo by targeted insertion of a novel IRES-Cre cassette into endogenous loci. *Mech Dev* 1999. **85**: 35-47.

408 **Hartzog, G. A., Speer, J. L. and Lindstrom, D. L.,** Transcript elongation on a nucleoprotein template. *Biochim Biophys Acta* 2002. **1577**: 276-286.

409 **Wolfer, A., Bakker, T., Wilson, A., Nicolas, M., Ioannidis, V., Littman, D. R., Lee, P. P., Wilson, C. B., Held, W., MacDonald, H. R. and Radtke, F.,** Inactivation of Notch 1 in immature thymocytes does not perturb CD4 or CD8T cell development. *Nat Immunol* 2001. **2**: 235-241.

410 **Kuhn, R., Schwenk, F., Aguet, M. and Rajewsky, K.,** Inducible gene targeting in mice. *Science* 1995. **269**: 1427-1429.
411 **Rickert, R. C., Roes, J. and Rajewsky, K.,** B lymphocyte-specific, Cre-mediated mutagenesis in mice. *Nucleic Acids Res* 1997. **25**: 1317-1318.
412 **Brocard, J., Feil, R., Chambon, P. and Metzger, D.,** A chimeric Cre recombinase inducible by synthetic,but not by natural ligands of the glucocorticoid receptor. *Nucleic Acids Res* 1998. **26**: 4086-4090.
413 **Kellendonk, C., Tronche, F., Casanova, E., Anlag, K., Opherk, C. and Schutz, G.,** Inducible site-specific recombination in the brain. *J Mol Biol* 1999. **285**: 175-182.
414 **Dietrich, P., Dragatsis, I., Xuan, S., Zeitlin, S. and Efstratiadis, A.,** Conditional mutagenesis in mice with heat shock promoter-driven cre transgenes. *Mamm Genome* 2000. **11**: 196-205.
415 **Zhang, Y., Riesterer, C., Ayrall, A. M., Sablitzky, F., Littlewood, T. D. and Reth, M.,** Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res* 1996. **24**: 543-548.
416 **Weber, P., Metzger, D. and Chambon, P.,** Temporally controlled targeted somatic mutagenesis in the mouse brain. *Eur J Neurosci* 2001. **14**: 1777-1783.
417 **Le Hir, H., Nott, A. and Moore, M. J.,** How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 2003. **28**: 215-220.
418 **Garrick, D., Fiering, S., Martin, D. I. and Whitelaw, E.,** Repeat-induced gene silencing in mammals. *Nat Genet* 1998. **18**: 56-59.
419 **Davis, B. P. and MacDonald, R. J.,** Limited transcription of rat elastase I transgene repeats in transgenic mice. *Genes Dev* 1988. **2**: 13-22.
420 **Linn, F., Heidmann, I., Saedler, H. and Meyer, P.,** Epigenetic changes in the expression of the maize A1 gene in Petunia hybrida: role of numbers of integrated gene copies and state of methylation. *Mol Gen Genet* 1990. **222**: 329-336.
421 **Assaad, F. F., Tucker, K. L. and Signer, E. R.,** Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. *Plant Mol Biol* 1993. **22**: 1067-1085.
422 **Sharpe, J. A., Wells, D. J., Whitelaw, E., Vyas, P., Higgs, D. R. and Wood, W. G.,** Analysis of the human alpha-globin gene cluster in transgenic mice. *Proc Natl Acad Sci U S A* 1993. **90**: 11262-11266.
423 **Scheid, O. M., Paszkowski, J. and Potrykus, I.,** Reversible inactivation of a transgene in Arabidopsis thaliana. *Mol Gen Genet* 1991. **228**: 104-112.
424 **O'Gorman, S., Dagenais, N. A., Qian, M. and Marchuk, Y.,** Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. *Proc Natl Acad Sci U S A* 1997. **94**: 14602-14607.
425 **da Rocha, S. T. and Ferguson-Smith, A. C.,** Genomic imprinting. *Curr Biol* 2004. **14**: R646-649.
426 **Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R.,** DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* 2003. **422**: 909-913.
427 **Walmsley, M. J., Ooi, S. K., Reynolds, L. F., Smith, S. H., Ruf, S., Mathiot, A., Vanes, L., Williams, D. A., Cancro, M. P. and Tybulewicz, V. L.,** Critical roles for Rac1 and Rac2 GTPases in B cell development and signaling. *Science* 2003. **302**: 459-462.
428 **Montoliu, L., Chavez, S. and Vidal, M.,** Variegation associated with lacZ in transgenic animals: a warning note. *Transgenic Res* 2000. **9**: 237-239.

429 **Chevalier-Mariette, C., Henry, I., Montfort, L., Capgras, S., Forlani, S., Muschler, J. and Nicolas, J. F.,** CpG content affects gene silencing in mice: evidence from novel transgenes. *Genome Biol* 2003. **4**: R53.

430 **Guy, L. G., Kothary, R., DeRepentigny, Y., Delvoye, N., Ellis, J. and Wall, L.,** The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. *Embo J* 1996. **15**: 3713-3721.

431 **Guy, L. G., Kothary, R. and Wall, L.,** Position effects in mice carrying a lacZ transgene in cis with the beta-globin LCR can be explained by a graded model. *Nucleic Acids Res* 1997. **25**: 4400-4407.

432 **Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. and Razin, A.,** The absence of detectable methylated bases in Drosophila melanogaster DNA. *FEBS Lett* 1982. **146**: 148-152.

433 **Lu, B. Y., Bishop, C. P. and Eissenberg, J. C.,** Developmental timing and tissue specificity of heterochromatin-mediated silencing. *Embo J* 1996. **15**: 1323-1332.

434 **Lyko, F., Ramsahoye, B. H. and Jaenisch, R.,** DNA methylation in Drosophila melanogaster. *Nature* 2000. **408**: 538-540.

435 **Razin, A.,** CpG methylation, chromatin structure and gene silencing-a three-way connection. *Embo J* 1998. **17**: 4905-4908.

436 **Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. and Li, E.,** Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* 2003. **23**: 5594-5605.

437 **Morgan, H. D., Santos, F., Green, K., Dean, W. and Reik, W.,** Epigenetic reprogramming in mammals. *Hum Mol Genet* 2005. **14 Spec No 1**: R47-58.

438 **Chaillet, J. R., Bader, D. S. and Leder, P.,** Regulation of genomic imprinting by gametic and embryonic processes. *Genes Dev* 1995. **9**: 1177-1187.

439 **Chaillet, J. R., Vogt, T. F., Beier, D. R. and Leder, P.,** Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell* 1991. **66**: 77-83.

440 **Kearns, M., Preis, J., McDonald, M., Morris, C. and Whitelaw, E.,** Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. *Nucleic Acids Res* 2000. **28**: 3301-3309.

441 **Preis, J. I., Downes, M., Oates, N. A., Rasko, J. E. and Whitelaw, E.,** Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. *Curr Biol* 2003. **13**: 955-959.

442 **Shimshek, D. R., Kim, J., Hubner, M. R., Spergel, D. J., Buchholz, F., Casanova, E., Stewart, A. F., Seeburg, P. H. and Sprengel, R.,** Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* 2002. **32**: 19-26.

443 **Tybulewicz, V. L., Ardouin, L., Prisco, A. and Reynolds, L. F.,** Vav1: a key signal transducer downstream of the TCR. *Immunol Rev* 2003. **192**: 42-52.

444 **Ogilvy, S., Metcalf, D., Gibson, L., Bath, M. L., Harris, A. W. and Adams, J. M.,** Promoter elements of vav drive transgene expression in vivo throughout the hematopoietic compartment. *Blood* 1999. **94**: 1855-1863.

445 **Zhumabekov, T., Corbella, P., Tolaini, M. and Kioussis, D.,** Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. *J Immunol Methods* 1995. **185**: 133-140.

446 **Srinivas, S., Watanabe, T., Lin, C. S., William, C. M., Tanabe, Y., Jessell, T. M. and Costantini, F.,** Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* 2001. **1**: 4.

447 **Singbartl, K., Thatte, J., Smith, M. L., Wethmar, K., Day, K. and Ley, K.,** A CD2-green fluorescence protein-transgenic mouse reveals very late antigen-4-dependent CD8+ lymphocyte rolling in inflamed venules. *J Immunol* 2001. **166**: 7520-7526.

448 **Godfrey, D. I. and Zlotnik, A.,** Control points in early T-cell development. *Immunol Today* 1993. **14**: 547-553.

449 **Shortman, K.,** Cellular aspects of early T-cell development. *Curr Opin Immunol* 1992. **4**: 140-146.

450 **de Boer, J., Williams, A., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A. J. and Kioussis, D.,** Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur J Immunol* 2003. **33**: 314-325.

451 **Okumura, K., Kaneko, Y., Nonoguchi, K., Nishiyama, H., Yokoi, H., Higuchi, T., Itoh, K., Yoshida, O., Miki, T. and Fujita, J.,** Expression of a novel isoform of Vav, Vav-T, containing a single Src homology 3 domain in murine testicular germ cells. *Oncogene* 1997. **14**: 713-720.

452 **Cui, C., Wani, M. A., Wight, D., Kopchick, J. and Stambrook, P. J.,** Reporter genes in transgenic mice. *Transgenic Res* 1994. **3**: 182-194.

453 **Kelley, K. A., Friedrich, V. L., Jr., Sonshine, A., Hu, Y., Lax, J., Li, J., Drinkwater, D., Dressler, H. and Herrup, K.,** Expression of Thy-1/lacZ fusion genes in the CNS of transgenic mice. *Brain Res Mol Brain Res* 1994. **24**: 261-274.

454 **Eckardt, D., Theis, M., Doring, B., Speidel, D., Willecke, K. and Ott, T.,** Spontaneous ectopic recombination in cell-type-specific Cre mice removes loxP-flanked marker cassettes in vivo. *Genesis* 2004. **38**: 159-165.

455 **Opsahl, M. L., McClenaghan, M., Springbett, A., Reid, S., Lathe, R., Colman, A. and Whitelaw, C. B.,** Multiple effects of genetic background on variegated transgene expression in mice. *Genetics* 2002. **160**: 1107-1112.

456 **Georgiades, P., Ogilvy, S., Duval, H., Licence, D. R., Charnock-Jones, D. S., Smith, S. K. and Print, C. G.,** vavCre Transgenic mice: A tool for mutagenesis in hematopoietic and endothelial lineages. *Genesis* 2002. **34**: 251-256.

457 **Ogilvy, S., Elefanty, A. G., Visvader, J., Bath, M. L., Harris, A. W. and Adams, J. M.,** Transcriptional regulation of vav, a gene expressed throughout the hematopoietic compartment. *Blood* 1998. **91**: 419-430.

458 **Wilson, T. J., Cowdery, H. E., Xu, D., Kola, I. and Hertzog, P. J.,** A human CD2 minigene directs CRE-mediated recombination in T cells in vivo. *Genesis* 2002. **33**: 181-184.

459 **Loonstra, A., Vooijs, M., Beverloo, H. B., Allak, B. A., van Drunen, E., Kanaar, R., Berns, A. and Jonkers, J.,** Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A* 2001. **98**: 9209-9214.

460 **Strebel, A., Harr, T., Bachmann, F., Wernli, M. and Erb, P.,** Green fluorescent protein as a novel tool to measure apoptosis and necrosis. *Cytometry* 2001. **43**: 126-133.

461 **Steff, A. M., Fortin, M., Arguin, C. and Hugo, P.,** Detection of a decrease in green fluorescent protein fluorescence for the monitoring of cell death: an assay

amenable to high-throughput screening technologies. *Cytometry* 2001. **45**: 237-243.

462 **Collier, R. J.,** Diphtheria toxin: mode of action and structure. *Bacteriol Rev* 1975. **39**: 54-85.

463 **Spilsberg, B., Hanada, K. and Sandvig, K.,** Diphtheria toxin translocation across cellular membranes is regulated by sphingolipids. *Biochem Biophys Res Commun* 2005. **329**: 465-473.

464 **Collier, R. J.,** Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. *Toxicon* 2001. **39**: 1793-1803.

465 **Yamaizumi, M., Mekada, E., Uchida, T. and Okada, Y.,** One molecule of diphtheria toxin fragment A introduced into a cell can kill the cell. *Cell* 1978. **15**: 245-250.

466 **Palmiter, R. D., Behringer, R. R., Quaife, C. J., Maxwell, F., Maxwell, I. H. and Brinster, R. L.,** Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. *Cell* 1987. **50**: 435-443.

467 **Breitman, M. L., Clapoff, S., Rossant, J., Tsui, L. C., Glode, L. M., Maxwell, I. H. and Bernstein, A.,** Genetic ablation: targeted expression of a toxin gene causes microphthalmia in transgenic mice. *Science* 1987. **238**: 1563-1565.

468 **Breitman, M. L., Rombola, H., Maxwell, I. H., Klintworth, G. K. and Bernstein, A.,** Genetic ablation in transgenic mice with an attenuated diphtheria toxin A gene. *Mol Cell Biol* 1990. **10**: 474-479.

469 **Breitman, M. L., Bryce, D. M., Giddens, E., Clapoff, S., Goring, D., Tsui, L. C., Klintworth, G. K. and Bernstein, A.,** Analysis of lens cell fate and eye morphogenesis in transgenic mice ablated for cells of the lens lineage. *Development* 1989. **106**: 457-463.

470 **Behringer, R. R., Mathews, L. S., Palmiter, R. D. and Brinster, R. L.,** Dwarf mice produced by genetic ablation of growth hormone-expressing cells. *Genes Dev* 1988. **2**: 453-461.

471 **Lee, K. J., Dietrich, P. and Jessell, T. M.,** Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. *Nature* 2000. **403**: 734-740.

472 **Lee, P., Morley, G., Huang, Q., Fischer, A., Seiler, S., Horner, J. W., Factor, S., Vaidya, D., Jalife, J. and Fishman, G. I.,** Conditional lineage ablation to model human diseases. *Proc Natl Acad Sci U S A* 1998. **95**: 11371-11376.

473 **Bartell, J. G., Fantz, D. A., Davis, T., Dewey, M. J., Kistler, M. K. and Kistler, W. S.,** Elimination of male germ cells in transgenic mice by the diphtheria toxin A chain gene directed by the histone H1t promoter. *Biol Reprod* 2000. **63**: 409-416.

474 **Matsumura, H., Hasuwa, H., Inoue, N., Ikawa, M. and Okabe, M.,** Lineage-specific cell disruption in living mice by Cre-mediated expression of diphtheria toxin A chain. *Biochem Biophys Res Commun* 2004. **321**: 275-279.

475 **Brockschnieder, D., Lappe-Siefke, C., Goebbels, S., Boesl, M. R., Nave, K. A. and Riethmacher, D.,** Cell depletion due to diphtheria toxin fragment A after Cre-mediated recombination. *Mol Cell Biol* 2004. **24**: 7636-7642.

476 **Sato, M. and Tanigawa, M.,** Production of CETD transgenic mouse line allowing ablation of any type of specific cell population. *Mol Reprod Dev* 2005. **72**: 54-67.

477 **Grieshammer, U., Lewandoski, M., Prevette, D., Oppenheim, R. W. and Martin, G. R.,** Muscle-specific cell ablation conditional upon Cre-mediated

DNA recombination in transgenic mice leads to massive spinal and cranial motoneuron loss. *Dev Biol* 1998. **197**: 234-247.

478 **Hardy, R. R. and Hayakawa, K.,** B cell development pathways. *Annu Rev Immunol* 2001. **19**: 595-621.

479 **Inui, S., Maeda, K., Hua, D. R., Yamashita, T., Yamamoto, H., Miyamoto, E., Aizawa, S. and Sakaguchi, N.,** BCR signal through alpha 4 is involved in S6 kinase activation and required for B cell maturation including isotype switching and V region somatic hypermutation. *Int Immunol* 2002. **14**: 177-187.

480 **Horcher, M., Souabni, A. and Busslinger, M.,** Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity* 2001. **14**: 779-790.

481 **Pham, C. T., MacIvor, D. M., Hug, B. A., Heusel, J. W. and Ley, T. J.,** Long-range disruption of gene expression by a selectable marker cassette. *Proc Natl Acad Sci U S A* 1996. **93**: 13090-13095.

482 **Mackall, C. L., Hakim, F. T. and Gress, R. E.,** Restoration of T-cell homeostasis after T-cell depletion. *Semin Immunol* 1997. **9**: 339-346.

483 **Rocha, B., Dautigny, N. and Pereira, P.,** Peripheral T lymphocytes: expansion potential and homeostatic regulation of pool sizes and CD4/CD8 ratios in vivo. *Eur J Immunol* 1989. **19**: 905-911.

484 **Agenes, F. and Freitas, A. A.,** Transfer of small resting B cells into immunodeficient hosts results in the selection of a self-renewing activated B cell population. *J Exp Med* 1999. **189**: 319-330.

485 **Gaudin, E., Rosado, M., Agenes, F., McLean, A. and Freitas, A. A.,** B-cell homeostasis, competition, resources, and positive selection by self-antigens. *Immunol Rev* 2004. **197**: 102-115.

486 **Ge, Q., Hu, H., Eisen, H. N. and Chen, J.,** Different contributions of thymopoiesis and homeostasis-driven proliferation to the reconstitution of naive and memory T cell compartments. *Proc Natl Acad Sci U S A* 2002. **99**: 2989-2994.

487 **McKnight, R. A., Shamay, A., Sankaran, L., Wall, R. J. and Hennighausen, L.,** Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proc Natl Acad Sci U S A* 1992. **89**: 6943-6947.

488 **Phi-Van, L. and Stratling, W. H.,** Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry* 1996. **35**: 10735-10742.

489 **Zhang, S. B. and Qian, R. L.,** The interaction between the human beta-globin locus control region and nuclear matrix. *Cell Res* 2002. **12**: 411-416.

490 **Loc, P. V. and Stratling, W. H.,** The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *Embo J* 1988. **7**: 655-664.

491 **Wotton, D., Flanagan, B. F. and Owen, M. J.,** Chromatin configuration of the human CD2 gene locus during T-cell development. *Proc Natl Acad Sci U S A* 1989. **86**: 4195-4199.

492 **Kowolik, C. M., Hu, J. and Yee, J. K.,** Locus control region of the human CD2 gene in a lentivirus vector confers position-independent transgene expression. *J Virol* 2001. **75**: 4641-4648.

493 **Indraccolo, S., Minuzzo, S., Roccaforte, F., Zamarchi, R., Habeler, W., Stievano, L., Tosello, V., Klein, D., Gunzburg, W. H., Basso, G., Chieco-Bianchi, L. and Amadori, A.,** Effects of CD2 locus control region sequences

on gene expression by retroviral and lentiviral vectors. *Blood* 2001. **98**: 3607-3617.

494 **Rank, G., Prestel, M. and Paro, R.,** Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. *Mol Cell Biol* 2002. **22**: 8026-8034.

495 **Travers, A.,** Chromatin modification by DNA tracking. *Proc Natl Acad Sci U S A* 1999. **96**: 13634-13637.

496 **Bolland, D. J., Wood, A. L., Johnston, C. M., Bunting, S. F., Morgan, G., Chakalova, L., Fraser, P. J. and Corcoran, A. E.,** Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 2004. **5**: 630-637.

497 **Rogan, D. F., Cousins, D. J. and Staynov, D. Z.,** Intergenic transcription occurs throughout the human IL-4/IL-13 gene cluster. *Biochem Biophys Res Commun* 1999. **255**: 556-561.

498 **Rogan, D. F., Cousins, D. J., Santangelo, S., Ioannou, P. A., Antoniou, M., Lee, T. H. and Staynov, D. Z.,** Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci U S A* 2004. **101**: 2446-2451.

499 **Masternak, K., Peyraud, N., Krawczyk, M., Barras, E. and Reith, W.,** Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* 2003. **4**: 132-137.

500 **Batzer, M. A. and Deininger, P. L.,** Alu repeats and human genomic diversity. *Nat Rev Genet* 2002. **3**: 370-379.

501 **Stancheva, I.,** Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. *Biochem Cell Biol* 2005. **83**: 385-395.

502 **Volpe, P.,** The language of methylation in genomics of eukaryotes. *Biochemistry (Mosc)* 2005. **70**: 584-595.

503 **Kress, C., Thomassin, H. and Grange, T.,** Local DNA demethylation in vertebrates: how could it be performed and targeted? *FEBS Lett* 2001. **494**: 135-140.

504 **Diaz, P., Cado, D. and Winoto, A.,** A locus control region in the T cell receptor alpha/delta locus. *Immunity* 1994. **1**: 207-217.

505 **Ortiz, B. D., Cado, D., Chen, V., Diaz, P. W. and Winoto, A.,** Adjacent DNA elements dominantly restrict the ubiquitous activity of a novel chromatin-opening region to specific tissues. *Embo J* 1997. **16**: 5037-5045.

506 **Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K. and Brahmachari, S. K.,** Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. *Bioinformatics* 2004. **20**: 813-817.

507 **Li, T. H., Kim, C., Rubin, C. M. and Schmid, C. W.,** K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. *Nucleic Acids Res* 2000. **28**: 3031-3039.

508 **Liu, W. M., Maraia, R. J., Rubin, C. M. and Schmid, C. W.,** Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res* 1994. **22**: 1087-1095.

509 **Willoughby, D. A., Vilalta, A. and Oshima, R. G.,** An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J Biol Chem* 2000. **275**: 759-768.

510 **Thorey, I. S., Cecena, G., Reynolds, W. and Oshima, R. G.,** Alu sequence involvement in transcriptional insulation of the keratin 18 gene in transgenic mice. *Mol Cell Biol* 1993. **13**: 6742-6751.

511 **Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T. and Bergman, Y.,** A role for nuclear NF-kappaB in B-cell-specific demethylation of the Igkappa locus. *Nat Genet* 1996. **13**: 435-441.

expression was seen on T cells derived from EYFP-single transgenic controls (Fig. 3, 4 dotted line).

To determine precisely the time point at which Cre expression is initiated during thymocyte development in each of the lines, double-negative (DN) thymocytes were stained with anti-CD44 and anti-CD25 in order to separate the four stages of DN thymocyte development (Fig. 5) [39, 40]. When thymocytes enter the thymus they express high levels of CD44, but no CD25 (DN1 stage). As the thymocytes mature they up-regulate CD25 to become $CD44^+/CD25^+$ cells (DN2 stage); subsequent